

Université de Sherbrooke

**Identification de rétropseudogènes dérivés des ARNs hY et étude de leur fonction
dans l'épissage alternatif chez l'humain et Recherche d'un homologue de la
protéine RoBP1 chez la levure *Saccharomyces cerevisiae***

Par
Jean-François Noël
Département de Microbiologie et Infectiologie

Mémoire présenté à la Faculté de Médecine
en vue de l'obtention du grade de
maître ès sciences (M.Sc.) en Microbiologie

13 juin 2006



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-18993-1

Our file Notre référence

ISBN: 978-0-494-18993-1

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

RÉSUMÉ

Identification de rétropseudogènes dérivés des ARNs hY et étude de leur fonction dans l'épissage alternatif chez l'humain et Recherche d'un homologue de la protéine RoBP1 chez la levure *Saccharomyces cerevisiae*

Par Jean-François Noël

Département de Microbiologie et Infectiologie, Université de Sherbrooke.

Mémoire présenté à la Faculté de Médecine en vue de l'obtention du grade de maître ès sciences (M.Sc.) en Microbiologie le 13 juin 2006.

Les ribonucléoprotéines Ro humaines sont constituées d'un des 4 ARNs hY et des protéines Ro60 et La, et probablement des protéines RoBP1, PTB et hnRNP K. Certaines maladies autoimmunes affectant les tissus conjonctifs sont caractérisées par la formation d'autoanticorps contre ces ribonucléoprotéines. Aucune fonction n'est attribuée aux RNPs Ro et leur rôle dans la pathogénèse des maladies qui sont associées aux anticorps ciblant ces RNPs reste indéfini.

La protéine RoBP1 interagit spécifiquement avec les RNPs Ro contenant Ro60 et hY5. Les particularités biochimiques et immunologiques de ces RNPs Ro suggèrent une fonction particulière qui pourrait être découverte par une meilleure connaissance de la protéine RoBP1. Des travaux antérieurs tentant d'identifier

l'homologue fonctionnel de RoBP1 chez la levure suggéraient l'existence de 2 gènes pouvant compléter pour la perte de RoBP1 dans une souche de levure rendue dépendante de RoBP1. Nous avons confirmé la dépendance à RoBP1 de cette souche. Cependant, les petites colonies sur milieu riche et l'absence de croissance sur milieu non-fermentable nous ont indiqué des déficiences respiratoires. Un rétrocroisement de cette souche avec une levure sauvage ne nous a pas permis de corriger ces déficiences. De plus, les tests de complémentation non concluants et l'absence de létalité des délétions des 2 gènes (UBC5 et CAF4) contredisent les travaux antérieurs. Donc, aucun homologue fonctionnel de RoBP1 n'est identifié chez la levure.

Les gènes codants pour les ARNs hY semblent être présents en une seule copie par génome. Cependant, il existe une abondance de séquences homologues aux ARNs hY dans le génome humain. Les séquences répétées les plus représentées dans le génome proviennent des éléments transposables, dont le rétrotransposon L1. La machinerie de rétrotransposition de l'élément L1 peut être utilisée pour la mobilisation d'autres séquences, comme les éléments Alu, et cette rétrotransposition laisse des marques caractéristiques identifiables dans le génome. L'analyse approfondie des séquences hY génomiques pourrait permettre de découvrir leur origine. À l'aide d'outils bioinformatiques, nous avons identifié 966 pseudogènes dérivés des ARNs hY dispersés dans tous les chromosomes avec une préférence pour les régions riches en gènes. La conservation des séquences et la distribution génomique des pseudogènes hY sont similaires à ceux des éléments Alu et leur apparition serait survenue après la divergence des rongeurs et des primates. L'analyse des séquences nous a permis d'identifier plusieurs caractéristiques suggérant que les

pseudogènes hY ont été mobilisés directement par la machinerie de rétrotransposition de L1: queue poly(A), duplications du site cible flanquant le pseudogène, et site d'insertion similaire au site consensus reconnu par l'endonuclease de l'élément L1. Nous avons identifié chez les pseudogènes hY certaines mutations ponctuelles spécifiques au niveau des sites de liaison aux protéines Ro60, La ou hnRNP K.

Trois des protéines interagissant avec les RNPs Ro seraient impliquées dans l'épissage (PTB, hnRNP K et RoBP1). Des séquences homologues aux ARN hY se retrouvant dans des introns près de sites d'épissage pourraient recruter ces protéines et moduler l'épissage de l'intron. Nous avons introduit des séquences correspondant aux ARNs hY dans les introns de modèles d'épissage alternatif (hnRNP A1 et Bcl-x) pour vérifier leur impact sur l'épissage. Nous avons observé qu'une séquence hY3 en aval ou en amont de l'exon alternatif du modèle A1 favorise l'exclusion de cet exon; modulation qui est plus importante lorsque la séquence est dans l'orientation opposée au gène. Des expériences de compétition montrent cependant que cette modulation semble être indépendante de l'action d'un facteur agissant en *trans*. Dans certains cas, la modulation observée est atténuée par la présence dans le mélange d'épissage d'une séquence complémentaire. Contrairement au modèle A1, l'introduction de séquences hY3 introniques influence peu l'épissage *in vitro* et *in vivo* du mini-gène Bcl-x.

Mots clés : ARN hY; RoBP1; pseudogène; épissage; rétrotransposition.

TABLE DES MATIÈRES

<u>TABLE DES MATIÈRES</u>	i
<u>LISTE DES FIGURES</u>	vii
<u>LISTE DES TABLEAUX</u>	ix
<u>LISTE DES ABRÉVIATIONS</u>	x
<u>RÉSUMÉ</u>	xv
<u>1. INTRODUCTION</u>	1
1.1 Les RNPs Ro.....	1
1.1.1 Les ARNs Y.....	1
1.1.2 Les protéines associées aux RNPs Ro.....	4
1.1.2.1 La protéine Ro60.....	4
1.1.2.2 La protéine La.....	6
1.1.2.3 La protéine RoBP1.....	8
1.1.2.4 Les autres protéines.....	10
1.1.3 Les RNPs Ro et l'autoimmunité.....	15
1.2 Homologues fonctionnels de RoBP1 chez la levure <i>Saccharomyces cerevisiae</i> ...16	
1.2.1 Création d'une levure <i>S. cerevisiae</i> dépendante de RoBP1.....	17
1.2.2 Tests de complémentation pour trouver l'homologue de RoBP1.....	18
1.3 Les pseudogènes hY.....	19
1.3.1 L'abondance de séquences génomiques homologues aux ARNs Y.....	19
1.3.2 Les éléments transposables dans le génome.....	21

1.3.3. L'élément L1.....	26
1.3.4 L'élément Alu.....	28
1.3.5 Le mécanisme de rétrotransposition des LINEs et des SINEs.....	29
1.3.6 L'influence des éléments mobiles.....	32
1.4. Les RNPs Ro et l'épissage.....	35
1.4.1 L'épissage constitutif et alternatif.....	35
1.4.2 La régulation de l'épissage alternatif.....	38
1.4.2.1 Les éléments régulateurs exoniques.....	39
1.4.2.2 Les éléments régulateurs introniques.....	40
1.4.3 Les protéines associées à la RNP Ro et l'épissage.....	42
1.4.3.1 La protéine PTB.....	42
1.4.3.2 La protéine hnRNP K.....	44
1.4.3.3 La protéine RoBP1.....	45
2. <u>HYPOTHÈSES ET BUTS</u>	46
3. <u>MATÉRIEL ET MÉTHODES</u>	47
3.1. Manipulations de base de l'ADN.....	47
3.1.1 Culture bactérienne et transformation de plasmides.....	47
3.1.2 Purification de plasmides.....	47
3.1.3 Visualisation d'ADN sur gel d'agarose.....	48
3.1.4 Isolation de bandes d'ADN, digestion enzymatique et ligation.....	48
3.1.5 Amplifications par PCR.....	49
Levure et RoBP1	49
3.2 Souches de levure utilisées.....	49

3.3 Culture des levures.....	50
3.4 Transfert de levures d'un Pétri à un autre (replica plating).....	51
3.5 Vérification de la capacité respiratoire des levures.....	52
3.6 Rétrocroisement de la souche de levure dépendante de RoBP1.....	53
3.6.1 Croisement des levures.....	53
3.6.2 Identification des diploïdes.....	53
3.6.3 Sporulation des levures.....	54
3.6.4 Microdissection des spores.....	55
3.6.5 Analyse des spores.....	55
3.7 Construction des plasmides pRS313/CAF4 et pRS313/UBC5.....	57
3.8 Tests de complémentation.....	58
3.8.1 Plasmides utilisés pour les tests de complémentation.....	58
3.8.2 Transformation de plasmides dans les levures.....	59
3.8.3 Croissance sur milieu 5'FOA.....	60
3.9 Création de levures dont les gènes CAF4 et/ou UBC5 ont été supprimés.....	61
3.9.1 Construction des fragments de délétion.....	61
3.9.2 Transformation de fragments de délétion et sélection des clones positifs.....	62
3.9.3 Vérification des délétions.....	63
3.9.3.1 Préparation de l'ADN génomique de levures.....	63
3.9.3.2 Vérification des délétions par PCR.....	64
3.9.4 Sporulation, microdissection et analyse des spores.....	64
Pseudogènes hY.....	67

3.10 Recherche de séquences homologues.....	67
3.11 Analyse de variation de séquence.....	68
3.12 Contenu en GC.....	69
3.13 Localisation des pseudogènes.....	69
3.14 Analyse des signatures de rétrotransposition.....	69
3.15 Association avec les éléments Alu.....	71
hY et épissage.....	72
3.16. Construction des plasmides.....	72
3.16.1 Modèle d'épissage alternatif hnRNP A1.....	72
3.16.2 Modèle d'épissage alternatif Bcl-x.....	76
3.17. Transcription des ARNs pré-messagers.....	80
3.17.1 Préparation des matrices.....	80
3.17.1.1 Modèle hnRNP A1.....	80
3.17.1.2 Modèle Bcl-x.....	81
3.17.2 Réaction de transcription des ARNs pré-messagers.....	81
3.17.3 Purification et dosage des ARNs transcrits.....	82
3.18. Épissage des ARNs pré-messagers.....	83
3.18.1 Réaction d'épissage <i>in vitro</i>	83
3.18.2 Réaction d'épissage <i>in vitro</i> avec ARN hY compétiteur.....	84
3.18.3 Réaction d'épissage <i>in vivo</i> (Bcl-x).....	85
3.18.3.1 Transfection de cellules HeLa et ECR 293.....	85
3.18.3.2 Extraction d'ARN.....	85
3.19. Détection et quantification des produits d'épissage.....	85

3.19.1 Modèle hnRNP A1.....	85
3.19.2 Modèle Bcl-x.....	87
4. <u>RÉSULTATS</u>	88
RoBP1 et levure	88
4.1 Levure dépendante de RoBP1: confirmation de la dépendance.....	88
4.2 Capacité respiratoire de la souche dépendante de RoBP1 : la souche dépendante de RoBP1 a un phénotype de type petite.....	89
4.3 Analyse du phénotype de toutes les souches dépendantes de RoBP1.....	90
4.4 Rétrocroisement de la levure dépendante de RoBP1 avec une levure sauvage : la souche dépendante de RoBP1 reste petite.....	92
4.5 Tests de complémentation : aucune complémentation claire.....	94
4.6 Délétion des gènes UBC5 et CAF4 dans la levure BY4705.....	94
Pseudogènes hY	101
4.7 Identification de 966 pseudogènes des ARNs hY.....	101
4.8 Conservation et âge des pseudogènes hY.....	103
4.9 Localisation génomique des pseudogènes hY et statistiques.....	112
4.10 Rétrotransposition des ARN hY par la machinerie L1.....	113
4.11 Pseudogènes hY dupliqués.....	120
hY et épissage	121
4.12 Modèle d'épissage alternatif hnRNP A1.....	121
4.12.1 Les séquences hY3 introniques modulent l'épissage <i>in vitro</i>	122
4.12.1.1 Séquence hY3 en aval de l'exon alternatif.....	122
4.12.1.2 Séquence hY3 en amont de l'exon alternatif.....	124

4.12.2 Les effets sur l'épissage de la séquence hY3 intronique sont constants d'un extrait nucléaire à un autre.....	126
4.12.3 La modulation de l'épissage ne dépend pas d'un facteur <i>trans</i> : épissage avec ARN hY3 compétiteur.....	129
4.13 Modèle d'épissage alternatif Bcl-x.....	136
4.13.1 Les séquences hY3 introniques influencent peu l'épissage <i>in vitro</i>	137
4.13.2 Les séquences hY3 introniques influencent peu l'épissage <i>in vivo</i>	137
5. <u>DISCUSSION</u>	141
5.1 Levure et RoBP1.....	141
5.2 Pseudogènes hY.....	146
5.3 hY et épissage.....	149
5.3.1 Modèle d'épissage alternatif hnRNP A1.....	150
5.3.2 Recherche d'un facteur <i>trans</i>	153
5.3.3 Modèle d'épissage alternatif Bcl-x.....	156
<u>REMERCIEMENTS</u>	159
<u>RÉFÉRENCES</u>	161

LISTE DES FIGURES

Figure 1: Structures secondaires et séquences nucléotidiques des ARNs hY.....	3
Figure 2 : Structure des rétrotransposons L1 et Alu.....	27
Figure 3 : Mécanisme de rétrotransposition de L1 par TPRT.....	31
Figure 4 : Représentation de la méthode de microdissection des tétrades.....	56
Figure 5 : Vérification des délétions simples UBC5 et CAF4 par PCR.....	65
Figure 6 : Vérification de la délétion CAF4 dans la souche Δ UBC5 par PCR.....	66
Figure 7 : Schéma de l'unité d'épissage alternatif dérivée du gène hnRNP A1.....	74
Figure 8 : Schémas et nomenclature des constructions utilisées dans les réactions d'épissage <i>in vitro</i> du mini-gène hnRNP A1.....	75
Figure 9 : Schéma de l'unité d'épissage alternatif dérivée du gène Bcl-x.....	77
Figure 10 : Schémas et nomenclature des constructions utilisées dans les réactions d'épissage <i>in vitro</i> et <i>in vivo</i> du mini-gène Bcl-x.....	78
Figure 11 : La levure dépendante de RoBP1 donne de petites colonies sur milieu YEPD.....	91
Figure 12 : La levure dépendante de RoBP1 ne pousse pas sur milieu YEPG.....	91
Figure 13 : Vérification du remplacement du gène UBC5 par LEU2.....	96
Figure 14 : Vérification du remplacement du gène CAF4 par TRP1.....	97
Figure 15 : Vérification du remplacement du gène CAF4 par HIS3.....	98
Figure 16 : Distribution des pseudogènes hY dans le génome humain.....	104
Figure 17 : Corrélation entre le nombre de pseudogènes hY et la taille des	

chromosomes.....	105
Figure 18: Mutations les plus fréquentes dans les pseudogènes hY.....	108
Figure 19 : Conservation et âge des pseudogènes hY.....	109
Figure 20 : Exemples de pseudogènes rétrotransposés.....	117
Figure 21 : Modulation de l'épissage <i>in vitro</i> du mini-gène hnRNP A1 contenant une séquence hY3 en aval de l'exon alternatif.....	123
Figure 22 : Modulation de l'épissage <i>in vitro</i> du mini-gène hnRNP A1 contenant une séquence hY3 en amont de l'exon alternatif.....	125
Figure 23 : Effets constants de la séquence hY3 intronique sur l'épissage avec différents extraits nucléaires.....	128
Figure 24 : Modulation de l'épissage <i>in vitro</i> du mini-gène hnRNP A1 avec ARN hY3 compétiteur.....	131
Figure 25 : Modulation de l'épissage <i>in vitro</i> du mini-gène hnRNP A1 avec ARN hY3 antisens compétiteur.....	132
Figure 26: Modulation de l'épissage <i>in vitro</i> du mini-gène hnRNP A1 avec ARN hY3 ou ARN hY3 antisens compétiteur.....	134
Figure 27 : Épissage <i>in vitro</i> du mini-gène Bcl-x contenant des séquences hY3 introniques.....	138
Figure 28 : Épissage <i>in vivo</i> du mini-gène Bcl-x contenant des séquences hY3 introniques.....	140

LISTE DES TABLEAUX

Tableau 1 : Gènes présents sur les fragments d'ADN complétant la perte de RoBP1.....	20
Tableau 2 : Caractéristiques des différents types d'éléments transposables dans le génom humain.....	23
Tableau 3 : Phénotype des différentes souches de levure dépendantes de RoBP1 et des souches contrôles.....	93
Tableau 4 : Exemple de ségrégation des spores issues de la microdissection de la souche diploïde contenant la délétion du gène UBC5.....	100
Tableau 5 : Nombre de séquences homologues à différents petits ARNs non-codants dans le génom humain.....	102
Tableau 6 : Densité des pseudogènes hY dans le génom humain.....	106
Tableau 7 : Nombres relatif de pseudogènes Y chez différentes espèces.....	111
Tableau 8 : Statistiques globales des pseudogènes hY.....	114
Tableau 9 : Pseudogènes hY montrant des signatures de rétrotransposition par L1..	118

LISTE DES ABRÉVIATIONS

A ou Ade	adénine
amol	attomole (10^{-18})
APP	précurseur de la protéine amyloïde
ARNm	ARN messenger
ARNr	ARN ribosomal
ARNt	ARN de transfert
C	cytosine
Caf4	<i>CCR4-associated factor 4</i> ou facteur 4 associé au complexe CCR4
C-terminal	carboxy-terminal
D.O.	densité optique
dNTP	désoxyribonucléotide triphosphate
DTT	dithiothreitol
DNase	désoxyribonucléase
EDTA	acide éthylène diamine tétra-acétique
EMS	éthyl méthane sulfonate
ERV	rétrovirus endogène
ESE	<i>exonic splicing enhancer</i> ou élément exonique activateur de l'épissage
ESS	<i>exonic splicing silencer</i> ou élément exonique répresseur de l'épissage
FBP	<i>FUSE-binding protein</i> ou protéine liant l'élément FUSE
FIR	<i>FBP interacting repressor</i> ou répresseur interagissant avec FBP

5'FOA	acide 5'fluoroorotique
FUSE	<i>Far upstream element</i> ou élément éloigné en amont
G	guanine
HCl	acide chlorhydrique
His	histidine
hnRNP	RNP hétérogène nucléaire
IRES	<i>internal ribosome entry site</i> ou site interne d'entrée du ribosome
ISE	<i>intronic splicing enhancer</i> ou élément intronique activateur de l'épissage
ISS	<i>intronic splicing silencer</i> ou élément intronique répresseur de l'épissage
ITR	inversion terminale répétée
kb	kilobase
KCl	chlorure de potassium
kDa	kiloDalton
KH	<i>K homology</i> ou domaine d'homologie avec hnRNP K
L1	LINE 1
LB	Luria-Bertani
LED	lupus érythémateux disséminé
Leu	leucine
LiAc	acétate de lithium
LINE	<i>long interspersed nucleotide element</i> ou longue séquence d'ADN intercalée

LTR	longue répétition terminale
Lys	lysine
Mb	mégabase (10^6 bases)
Met	méthionine
MgCl ₂	chlorure de magnésium
NaOH	hydroxide de sodium
NCBI	<i>National Center for Biotechnology Information</i> ou Centre national d'informations biotechnologiques
NLS	signal de localisation nucléaire
NMD	<i>nonsense-mediated decay</i> ou dégradation des ARNm porteurs de codon non-sens
nt	nucléotide
N-terminal	amino-terminal
Pb	paire de bases
PCR	réaction de polymérisation en chaîne
PEG	polyéthylène glycol
PMSF	fluorure de phényl-méthyl sulfoxide
Pol	ARN polymérase
PTB	<i>polypyrimidine tract binding protein</i> ou protéine liant les séquences riches en pyrimdines
PUF60	<i>poly-U binding factor 60</i> ou facteur de 60 kDa liant les séquences riches en uridines
PUMP	PUF60, U2AF65, Mud2p interaction protéine–protéine

PVA	alcool polyvinyle
R	purine
RITA	<i>RNP Interaction Trap Assay</i> ou technique d'identification de protéines interagissant avec une RNP
RNase	ribonucléase
RNP	ribonucléoprotéine
rNTP	ribonucléotide triphosphate
RoBP1	<i>Ro RNP-binding protein 1</i> ou protéine 1 liant la RNP Ro
RRM	<i>RNA recognition motif</i> ou motif de liaison à l'ARN
SDS	sodium dodécyl sulfate
SC	<i>synthetic complete</i> ou milieu synthétique complet pour levure
SINE	<i>short interspersed nucleotide element</i> ou courte séquence d'ADN intercalée
SS	syndrome de Sjögren
snRNA	petit ARN nucléaire
snRNP	petite RNP nucléaire
T	thymine
TAE	tampon Tris-acétate-EDTA
TE	tampon Tris-EDTA
TFIIH	facteur de transcription IIH
TNE	tampon TRIS-NaOH-EDTA
TPRT	<i>target primed reverse transcription</i> ou rétrotranscription amorcée sur la cible

TRIM	motif tripartite
Tris	Tris (hydroxyméthyl) aminométhane
TROVE	module Télomérase, Ro et Vault
Trp	tryptophane
TSD	duplication du site cible
U	uridine
U2AF	facteur auxiliaire du snRNP U2
Ura	uracile
UTR	région non traduite
U.V.	rayons ultraviolets
V	volt
Y	pyrimidine
YNB	<i>yeast nitrogen base</i> ou base azotée de levure
Yc	<i>yeast complete</i> ou milieu synthétique complet pour levure
YEPD	<i>yeast extract peptone dextrose</i> ou milieu avec extrait de levure, peptone et dextrose
YEPG	<i>yeast extract peptone glycerol</i> ou milieu avec extrait de levure, peptone et glycérol

1. INTRODUCTION

1.1 Les RNPs Ro

Les RNPs Ro sont des complexes ribonucléoprotéiques de fonction inconnue et peu abondants, soit environ 10^5 copies par cellule (WOLIN & STEITZ, 1984). Ces petites particules, conservées dans l'évolution, sont retrouvées chez tous les mammifères testés. Des homologues des RNPs Ro sont également détectés dans une très grande variété d'organismes eucaryotes, dont le nématode *Caenorhabditis elegans* (VAN HORN et al., 1995) et la grenouille *Xenopus laevis* (O'BRIEN et al., 1993), et dans le procaryote *Deinococcus radiodurans* (CHEN et al., 2000). Ces particules n'ont cependant pas été détectées chez la levure. Les RNPs Ro sont composés d'un ARN de type Y, de la protéine Ro60 et, dans une certaine proportion, de la protéine La (HENDRICK et al., 1981; WOLIN & STEITZ, 1983; MAMULA et al., 1989). D'autres protéines peuvent également être associées à ce complexe. La localisation cellulaire des RNPs Ro est encore controversée : elles sont généralement reconnues comme étant cytoplasmiques (HENDRICK et al., 1981; PEEK et al., 1993), mais les RNPs Ro contenant l'ARN Y5 seraient plutôt nucléaires (GENDRON et al., 2001).

1.1.1 Les ARNs Y

Les ARNs Y sont de petits ARNs non-codant transcrits par l'ARN polymérase

(pol) III et leur nombre et taille varient en fonction des espèces et des types cellulaires (HENDRICK et al., 1981). Les ARN Y humains (hY), au nombre de quatre, sont nommés hY1, hY3, hY4 et hY5 (hY2 est un produit de dégradation de hY1) et ont une taille respective de 112, 101, 94 et 85 nucléotides (Figure 1; WOLIN & STEITZ, 1983). Les érythrocytes humains n'expriment que les ARN hY1 et hY4 (RADER et al., 1989), tandis que les plaquettes n'expriment que les ARN hY1 et hY3 (ITOHO & REICHLIN, 1991). La souris compte seulement 2 ARNs Y, nommés mY1 et mY3, qui sont homologues à 97% et 95% aux ARNs hY1 et hY3 respectivement. *C. elegans* et *X. leavis* possèdent respectivement un et quatre ARNs Y (VAN HORN et al., 1995; O'BRIEN et al., 1993). L'ARN Y3 est retrouvé chez le plus grand nombre d'organismes, tandis que l'ARN Y5 semble être présent seulement chez les mammifères supérieurs.

Bien que la plupart des molécules d'ARN Y sont associées à la protéine Ro60 dans le cytoplasme, quelques transcrits peuvent être détectés dans une structure périnucléolaire sans Ro60 (MATERA et al., 1995).

Les gènes des ARN hY sont présents en une seule copie dans le génome, contrairement à la plupart des ARNs transcrits par la pol III, et sont situés dans la même région du chromosome 7 (WOLIN & STEITZ, 1983; MARAIA et al., 1994; MARAIA et al., 1996). Les 2 gènes murins sont également uniques, mais très éloignés l'un de l'autre sur le chromosome 6 (FARRIS et al., 1996).

Bien que la séquence primaire des différents ARNs Y semble peu conservée, des études de comparaison phylogénétique et des expériences enzymatiques et biochimiques révèlent une structure secondaire en forme de tige-boucle, la tige étant

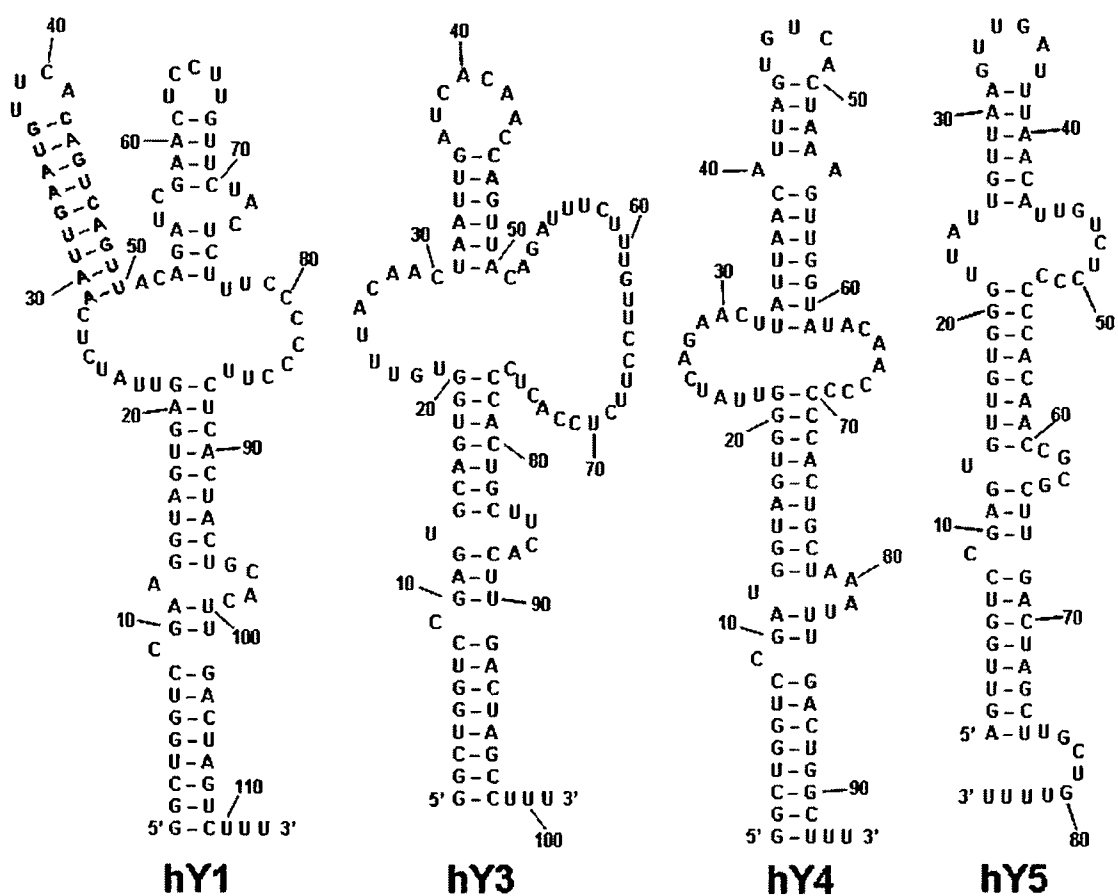


Figure 1: Structures secondaires et séquences nucléotidiques des ARNs hY.

formée de l'appariement des extrémités 5' et 3' (VAN GELDER et al., 1994; FARRIS et al., 1995; FARRIS et al., 1999; TEUNISSEN et al., 2000). Les séquences les mieux conservées sont situées dans cette tige et forment une hélice présentant une cytosine saillante. Cette région est responsable de la liaison de la protéine Ro60 (WOLIN & STEITZ, 1984; GREEN et al., 1998). La boucle interne est moins bien conservée et est souvent riche en pyrimidines. Les ARNs transcrits par la pol III présentent à leur extrémité 3' une courte série d'uridines qui est retirée au cours de la maturation (Figure 1). Par contre, chez les ARNs Y, cette queue polyuridylée est conservée et constitue le site de liaison de la protéine La (STEPHANO, 1984).

1.1.2 Les protéines associées aux RNPs Ro

1.1.2.1 La protéine Ro60

La protéine Ro60 est la composante protéique majeure des RNPs Ro. Cette protéine de faible abondance est présente dans toutes les cellules humaines et peut être localisée au cytoplasme et au noyau (PRUIJN et al., 1997). Ro60 est une protéine de 60 kDa qui présente 2 isoformes produits par épissage alternatif et contient un domaine de reconnaissance de l'ARN (RRM) et un motif en doigt de zinc (DEUTSCHER et al., 1988; BEN-CHETRIT et al., 1989). Le motif RRM est nécessaire, mais non suffisant pour la liaison aux ARN Y. De plus, la conformation de la protéine serait très importante pour cette interaction avec l'ARN (PRUIJN 1991). Par contre, des études bioinformatiques récentes attribuent plutôt la capacité

de liaison à l'ARN à un nouveau module N-terminal nommé TROVE. De plus, un domaine VWFA (von Willebrand Factor A) pourrait être impliqué dans des interactions protéine-protéine (BATEMAN & KICKHOEFER, 2003).

L'identification moléculaire de Ro60 chez l'humain, la souris, *X. leavis* et *C. elegans* montre que cette protéine est bien conservée dans l'évolution (DEUTSCHER et al., 1988; O'BRIEN et al., 1993; LABBE et al., 1995; VAN HORN et al., 1995; WANG et al., 1996). Par contre, le motif en doigt de zinc est une acquisition récente puisqu'il est seulement présent chez la protéine Ro60 humaine.

Dans les oocytes de *X. leavis*, la protéine Ro60 est associée à une classe variante de précurseurs de l'ARN ribosomal 5S. Ces pré-ARN 5S anormaux contiennent des nucléotides supplémentaires et des mutations causant un mauvais repliement de l'ARN (O'BRIEN & WOLIN, 1994; SHI et al., 1996). Dans les cellules souches embryonnaires de souris, Ro60 est aussi retrouvée associée à certains petits ARNs U2 anormalement repliés (CHEN et al., 2003). La protéine Ro60 pourrait donc participer au contrôle de qualité et à l'élimination des petits ARN défectueux.

La délétion de la protéine Ro60 chez *D. radiodurans*, une bactérie extrêmement résistante aux radiations, entraîne une diminution de la résistance aux radiations ultraviolettes (U.V.). De plus, dans la souche sauvage, le niveau de Ro60 augmente après exposition aux U.V. (CHEN et al., 2000). Dans les cellules de mammifère, la protéine Ro60 est aussi importante pour faciliter la survie cellulaire après exposition aux radiations ultraviolettes. De plus, Ro60 et un ARN Y s'accumulent au noyau suite à une irradiation aux U.V. (CHEN et al., 2003; XUE et

al., 2003). Donc Ro60 pourrait être impliquée dans la reconnaissance ou la réparation de dommages cellulaires causés par les rayons U.V. ou autres stress environnementaux.

Les souris déficientes en Ro60 développent également un syndrome autoimmun caractérisé par des autoanticorps contre la chromatine et les ribosomes, une glomérulonéphrite et une photosensibilité (XUE et al., 2003). Cette protéine pourrait donc être importante pour la prévention de maladies autoimmunes.

En collaboration avec la protéine La et un petit ARN inconnu, Ro60 pourrait aussi être impliquée dans la régulation de la traduction d'ARNs messagers codant pour certaines protéines ribosomales chez *X. leavis* (PELLIZONI et al., 1998).

1.1.2.2 La protéine La

La protéine La est une phosphoprotéine principalement nucléaire de 48 kDa qui est très abondante dans pratiquement tous les organismes et tous les types cellulaires (WOLIN & CEDERVALL, 2002). Cette protéine peut aussi se retrouver plus abondamment au cytoplasme dans certaines circonstances, comme après une infection virale (MEEROVITCH et al., 1993) et pendant l'apoptose (CASCIOLAROSSEN et al., 1994; AYUKAWA et al., 2000). La protéine La est associée transitoirement à tous les ARNs cellulaires nouvellement transcrits par la pol III au niveau de la séquence 3'-UUU_{OH}, mais demeure associée de façon stable avec les ARNs Y (BOIRE & CRAFT, 1990). La protéine La peut également lier plusieurs ARNs viraux (LERNER et al., 1981; CHANG et al., 1994; HEISE et al., 1999) et

certaines ARNs transcrits par l'ARN polymérase II dont l'extrémité 3' est uridylée (XUE et al., 2000; KUFEL et al., 2000).

L'analyse des homologues de la protéine La chez différentes espèces montre une très bonne conservation dans l'évolution. La portion N-terminale hautement conservée de la levure à l'humain contient le motif La qui pourrait participer à la spécificité de liaison à l'ARN. Le motif RRM1 est essentiel à la liaison à l'ARN et le motif RRM2 serait important pour une liaison de haute affinité (HORKE et al., 2004a). La portion C-terminale moins conservée contient un signal de localisation nucléaire (ROSENBLUM et al., 1998), un signal de rétention nucléaire (SIMONS et al., 1996), un domaine de dimérisation (CRAIG et al., 1997), un signal de localisation nucléolaire (HORKE et al., 2004b), un motif potentiel de liaison à l'ATP (TOPFER et al., 1993) et un site possible de clivage par des caspases (AYUKAWA et al., 2000).

La protéine La est une protéine multifonctionnelle. Elle a été longtemps décrite comme un facteur de terminaison de la transcription par la pol III (GOTTLIEB & STEITZ, 1989), mais les études plus récentes mettent en doute cette fonction (WESER et al., 2000). La protéine La serait plutôt impliquée dans la maturation des petits ARNs nouvellement transcrits, dont les ARNs de transfert et les petits ARN nucléolaires, en protégeant leur extrémité 3' contre une digestion par les exonucléases et en stabilisant leur structure secondaire (VAN HORN et al., 1997; YOO & WOLIN, 1997; PANNONE et al., 1998; XUE et al., 2000; KUFEL et al., 2000; CHAKSHUSMATHI et al., 2003). Une étude récente vient supporter cette activité chaperonne (BELISOVA et al., 2005).

La protéine La pourrait également contribuer à la rétention nucléaire de certains petits ARNs nouvellement transcrits, dont l'ARN hY1 (PEEK et al., 1993; SIMONS et al., 1996; GRIMM et al., 1997).

Plusieurs études montrent l'implication de la protéine La dans le processus de traduction. La protéine La régulerait la traduction CAP-indépendante (IRES-dépendante) d'ARNm viraux, comme le Coxsackievirus B3, le virus de l'hépatite C et le poliovirus (RAY & DAS, 2002; COSTA-MATTIOLI et al., 2004) et cellulaires (HOLCIK & KORNELUK, 2000; KIM et al., 2001). La protéine La pourrait aussi favoriser la traduction des ARNm ribosomaux contenant des séquences 5'-TOP (5' terminal oligopyrimidine), dans certains cas en collaboration avec Ro60 (PELLIZONI et al., 1998; CROSIO et al., 2000).

Finalement, La protéine La possède une activité hélicase ATP-dépendante permettant de séparer les ARNs double-brin (HUHN et al., 1997).

1.1.2.3 La protéine RoBP1

La protéine RoBP1 (Ro RNP-binding protein 1) fut identifiée en utilisant un système adapté du triple-hybride en levure, le système RITA (RNP interaction trap assay, BOUFFARD et al., 1999), comme interagissant spécifiquement avec les RNPs contenant la protéine Ro60 et l'ARN hY5. Cependant, cette interaction n'a pas encore été démontrée *in vivo*.

RoBP1 est une protéine de 60 kDa possédant 3 motifs RRM permettant la liaison à l'ARN et ayant une grande homologie avec le facteur d'épissage U2AF65

(BOUFFARD et al., 2000). Le troisième domaine RRM, divergeant d'un domaine RRM consensus, est un domaine PUMP participant aux interactions protéine-protéine pour former des dimères stables (PAGE-MCCAWE et al., 1999). RoBP1 est exprimée dans une grande variété de tissus et semble être surexprimée dans les gonades. RoBP1 a été localisée principalement dans le noyau.

Le gène codant pour RoBP1, aussi connu sous le nom de SIAHBP1, peut générer 3 isoformes qui diffèrent selon la présence ou l'absence de 2 courts exons alternatifs situés dans la région 5' (BOUFFARD et al., 2000). Les deux autres isoformes furent identifiées sous les noms de FIR (LIU et al., 2000) et PUF60 (PAGE-MCCAWE et al., 1999).

L'isoforme FIR (FBP interacting repressor) semble être impliquée dans la régulation de la transcription. La protéine FBP (FUSE-binding protein) lie l'élément FUSE (Far UpStream Element) situé en amont du promoteur pour activer la transcription du gène c-myc par TFIIH. La protéine FIR interfère avec FBP au niveau de l'élément FUSE et inhibe la portion hélicase de TFIIH via un domaine de répression N-terminal. FIR empêche donc la transcription de c-myc activée par FBP, mais non la transcription basale (LIU et al., 2000).

L'orthologue de PUF60/FIR/RoBP1 chez la Drosophile, Hfp, régule négativement la progression du cycle cellulaire durant le développement de la drosophile en inhibant l'expression de dMyc, ce qui est en accord avec les données concernant la protéine FIR chez les mammifères (QUINN et al., 2003). De plus, malgré sa localisation largement nucléaire, des évidences situent Hfp également au

niveau du cytoplasme et suggèrent un rôle dans la localisation cellulaire des ARN (VAN BUSKIRK & SCHÜPBACH, 2002).

Certaines études montrent également un rôle dans l'épissage, mais cette fonction sera abordée plus loin.

1.1.2.4 Les autres protéines

Plusieurs évidences suggèrent l'existence de sous-populations de RNPs Ro formées par l'association de différentes protéines supplémentaires (BOIRE & CRAFT, 1989; BOIRE & CRAFT, 1990; KELEKAR et al., 1994). Des expériences plus récentes de purification de RNPs Ro à partir d'extraits de cellules HeLa et de reconstitution de RNPs Ro avec des ARN Y biotinylés ont confirmé cette hypothèse (FABINI et al., 2000) et permis l'identification de la nucléoline (FOURAUX et al., 2002) et des protéines hnRNP K et PTB/hnRNP I (FABINI et al., 2001). Ces trois protéines lient spécifiquement les ARNs hY1 et hY3 au niveau de la boucle interne riche en pyrimidines. Cependant, les interactions détaillées entre ces protéines et l'ARN restent à définir.

PTB

La protéine PTB (pyrimidine tract binding)/hnRNP I fut découverte grâce à sa capacité à lier des séquences introniques riches en pyrimidines (GARCIA-BLANCO et al., 1989) et possède certaines caractéristiques la classant dans la famille des

protéines hnRNP (GHETTI et al., 1992). PTB est une abondante protéine de 58 kDa possédant 4 domaines de liaison à l'ARN du type RRM et un signal de localisation nucléaire (NLS) N-terminal. Les RRM2 et 3 seraient nécessaires à la dimérisation de la protéine et à la liaison de la protéine à l'ARN, respectivement (PEREZ et al., 1997; WAGNER & GARCIA-BLANCO, 2001). Le RRM4 serait important pour la fonction de la protéine, mais non pour sa liaison à l'ARN (LIU et al., 2002). Bien que localisée généralement au niveau du noyau, la protéine PTB peut également se retrouver au cytoplasme suite à une phosphorylation (XIE et al., 2003) ou à une digestion (BLACK et al., 2002). Trois isoformes de PTB générées par épissage alternatif ont été identifiées chez l'humain: PTB1, PTB2 et PTB4, et plusieurs homologues ont été observés chez d'autres organismes. La protéine PTB peut lier de courtes séquences d'ARN riches en pyrimidines, comme UUCU ou UCUCU, mais la liaison est optimale lorsque ces séquences sont situées dans un contexte aussi riche en pyrimidines (SINGH et al., 1995; PÉREZ et al., 1997).

Le rôle de PTB dans l'épissage est très bien documenté et sera abordé plus loin. En plus de son rôle dans l'épissage, PTB peut réguler négativement ou positivement la traduction indépendamment du CAP d'ARN viraux, comme le virus de l'hépatite C et un picornavirus (TISCHENDORF et al., 2004; PILIPENKO et al., 2001) et cellulaires, comme les protéines BAG-1 et BiP (PICKERING et al., 2003; KIM et al., 2000). PTB pourrait aussi localiser certains ARN messagers (COTE et al., 1999b) et moduler l'efficacité de polyadénylation (CASTELO-BRANCO et al., 2004).

hnRNP K

La protéine hnRNP K fut découverte initialement comme étant une des protéines associées aux ARN hétérogènes nucléaires (hnRNA) formant les particules hnRNP (MATUNIS et al., 1992). Quatre isoformes de hnRNP K produites par épissage alternatif ont été identifiées (hnRNP K A, B, C et D) et une cinquième pourrait correspondre à la protéine hnRNP J (DEJGAARD et al., 1994). La protéine hnRNP K possède un poids moléculaire apparent de 65 kDa et contient 3 domaines KH (K homology) qui sont très bien conservées dans l'évolution, de *Xenopus laevis* aux mammifères (SIOMI et al., 1993). Les domaines KH1 et KH2 sont groupés dans la partie N-terminale de la protéine et le domaine KH3 est situé en C-terminal. Ces domaines KH confèrent la spécificité de liaison de la protéine hnRNP K à de courtes séquences d'ARN simple-brin riches en C (SIOMI et al., 1994). De plus, les domaines KH sont impliqués dans l'affinité de la protéine pour l'ADN simple-brin et double-brin, mais pourraient aussi participer à la liaison avec des protéines. HnRNP K contient également un signal de localisation nucléaire (NLS), un domaine KNS (K nuclear shuttling) impliqué dans le transport de la protéine entre le noyau et le cytoplasme (MICHAEL et al., 1997), et une région d'interaction KI (K Interactive) qui est responsable des interactions protéines-protéines pouvant former des homodimères et des hétérodimères (KIM et al., 2000). Parmi les nombreux partenaires de hnRNP K, notons des protéines de signalisation cellulaire et des facteurs de transcription, d'épissage et de traduction (BOMSZTYK et al., 2004). La

phosphorylation de hnRNP K par une variété de kinases régule les interactions protéine-protéine et protéine-ARN (OSTROWSKI et al., 2000).

HnRNP K est une protéine multifonctionnelle impliquée dans une vaste gamme d'activités cellulaires : remodelage de la chromatine, activation et répression de la transcription, épissage, activation et répression de la traduction et stabilité de l'ARN (BOMSZTYK et al., 2004). HnRNP K pourrait agir comme une protéine d'échafaudage permettant le recrutement de différentes protéines effectrices.

Nucléoline

La nucléoline, appelée initialement C23, est la principale protéine nucléolaire des cellules en division active. La nucléoline est une phosphoprotéine hautement conservée dans l'évolution qui est impliquée dans de nombreux processus cellulaires. Cette protéine participe notamment à plusieurs étapes de la biogénèse des ribosomes, incluant la transcription des ADNs ribosomaux, la maturation des pré-ARNs ribosomaux et l'assemblage des ribosomes (GINISTY et al., 1999; SRIVASTAVA & POLLARD, 1999). La nucléoline possède quatre motifs RRM permettant la liaison aux ARNr (GHISOLFI-NIETO et al., 1996), un signal de localisation nucléaire (NLS), un domaine d'interaction protéine-protéine (RGG) responsable de l'interaction avec les protéines ribosomales (BOUVET et al., 2002) et un domaine N-terminal contenant des régions riches en résidus acides possiblement impliqué dans la régulation de la transcription des ADNs ribosomaux (ROGER et al., 2002). Transitant entre le nucléole et le cytoplasme, la nucléoline pourrait ainsi participer aux

transport des protéines ribosomales (BORER et al., 1989) et des ARNs Y (FOURAUX et al., 2002). La nucléoline serait aussi impliquée dans la prolifération et la croissance cellulaire ainsi que dans l'embryogenèse et la nucléogenèse (GINISTY et al., 1999; SRIVASTAVA & POLLARD, 1999). Finalement, la nucléoline pourrait jouer un rôle dans la traduction IRES-dépendante (IZUMI et al., 2001).

Ro52

Une autre protéine Ro de 52 kDa, la protéine Ro52, a été identifiée comme faisant partie du complexe RNP Ro par immunobuvardage avec des sérums reconnaissant les protéines Ro60 et La (BEN-CHETRIT et al., 1988; SLOBBE et al., 1991). Ro52 est une protéine peu abondante, principalement cytoplasmique, sans motif de liaison à l'ARN apparent, mais contenant un motif en doigt de zinc et un motif « leucine-zipper » qui permettraient la liaison à l'ADN (CHAN et al., 1991; ITOH et al., 1991; FRANK & MCCUBBIN, 1995; POURMAND et al., 1998; FRANK, 1999). Ro52 pourrait être impliquée dans l'activation des cellules T (ISHII et al., 2003), mais aucune fonction claire n'a encore été attribuée à Ro52. Des études récentes de bioinformatique et de génomique ont identifié Ro52 comme faisant partie d'une large famille de protéines à motif RING/Bbox/coiled-coil (RBCC), nommé aussi motif tripartite (TRIM). Les protéines de cette famille sont impliquées dans plusieurs processus cellulaires dont l'apoptose, la régulation du cycle cellulaire, la réponse virale et l'ubiquitination (MERONI & DIEZ-ROUX, 2005).

L'association de Ro52 avec les ARNs Y dépend de la présence de Ro60 et se ferait par interaction protéine-protéine avec Ro60 plutôt que directement avec l'ARN (PRUIJN et al., 1991; SLOBBE et al., 1992). Malgré tout, l'association de la protéine Ro52 avec les RNPs Ro reste controversée. Différentes expériences immunologiques utilisant des composantes des RNPs Ro purifiées biochimiquement ou par fractionnement sur gradient de glycérol montrent que la protéine Ro52 n'est pas une composante stable des RNPs Ro (KELEKAR et al., 1994; BOIRE et al., 1995).

Calréticuline

D'autres protéines sont également reconnues par des sérums reconnaissant les protéines Ro60 et La, mais leur association avec la RNP Ro est controversée. Parmi ces protéines, on retrouve la calréticuline (MCCAULIFFE et al., 1990; ROKEACH et al., 1991; CHENG et al., 1996). Cette protéine de 46 kDa, localisée principalement au réticulum endoplasmique, mais pouvant aussi se retrouver au noyau, est très bien conservée dans l'évolution. La calréticuline est une protéine multifonctionnelle impliquée dans l'homéostasie du calcium cellulaire, le contrôle de l'adhésion cellulaire, le chaperonnage moléculaire et la modulation de la régulation transcriptionnelle par des récepteurs hormonaux nucléaires (JOHNSON et al., 2001).

1.1.3 Les RNPs Ro et l'autoimmunité

Les RNPs Ro sont fréquemment la cible d'autoanticorps produits chez des

patients atteints de maladies du tissu conjonctif comme le syndrome de Sjögren (SS) et le lupus érythémateux disséminé (LED; TAN, 1989; HARLEY et al., 1992; YOUINOU et al., 1994). Des anticorps anti-Ro sont détectés dans 30 à 40% des cas de LED et dans 50 à 80% des cas de SS (HARLEY et al., 1986; REICHLIN, 1986). Des autoanticorps contre l'ADN, les histones et les snRNPs peuvent également être détectés. L'ARN hY5 est le seul des quatre ARNs hY contre lequel sont détectés des autoanticorps spécifiques (BOULANGER et al., 1995). L'implication des autoanticorps dans la pathogénèse des maladies autoimmunes qui leur sont associées reste à être définie.

1.2 Homologues fonctionnels de RoBP1 chez la levure *Saccharomyces cerevisiae*

Le rapport complexe entre l'antigénicité des RNPs Ro et le développement de maladies autoimmunes pourrait être mieux défini par une meilleure caractérisation des RNPs Ro. Les fonctions des RNPs Ro et de leurs partenaires cellulaires pourraient être directement responsables de leur immunogénicité. Le caractère particulier des RNPs Ro contenant l'ARN hY5 (localisation nucléaire, présence seulement chez les organismes supérieurs et antigénicité de l'ARN hY5) suggère une fonction particulière. Ainsi, l'étude de RoBP1, une protéine associée spécifiquement à cette classe distincte de RNPs Ro, pourrait aider à mieux comprendre cette fonction.

Une méthode éprouvée pour l'étude d'un gène ou d'une protéine humaine est d'étudier son homologue dans un autre organisme eucaryote plus simple, comme la levure. Bien qu'aucun homologue de séquence ne soit identifié pour RoBP1 chez la

levure, la recherche d'un homologue de fonction est possible en utilisant la mutagenèse et l'action de certains marqueurs de sélection (KRANZ & HOLM, 1990). Cette recherche a été amorcée par Jim Boulanger (laboratoire du Dr Gilles Boire) dans le cadre de ses travaux de maîtrise et est résumée ci-dessous.

1.2.1 Création d'une levure *S. cerevisiae* dépendante de RoBP1

Brièvement, un plasmide contenant le gène codant pour RoBP1 et les marqueurs URA3 et ADE2 a été transformé dans une souche de levure *S. cerevisiae*. Des mutations dans l'ADN ont été induites par un traitement des levures à l'éthyl méthane sulfonate (EMS). Les levures produisant des colonies blanches sur milieu riche et ne poussant pas sur milieu 5'FOA ont été récupérées. La couleur blanche indiquait la présence du marqueur ADE2. Le gène ADE2 code pour une enzyme impliquée dans la synthèse de l'adénine. L'absence de cette enzyme conduit à l'accumulation d'un composé qui, en condition aérobie, amène la formation d'un pigment rouge. Le milieu de culture 5'FOA est utilisé pour contre-sélectionner la présence du marqueur URA3 chez les levures. Le gène URA3 produit l'enzyme décarboxylase orotidine-5'phosphate requise pour la biosynthèse de l'uracile. Cette enzyme convertit l'acide 5'fluoroorotique (5'FOA) présent dans le milieu en 5-fluoruracile qui est un composé toxique pour la cellule (BOEKE et al., 1984). Ainsi, les levures contenant le marqueur URA3 situé sur le plasmide contenant RoBP1 pouvaient croître sur le milieu 5'FOA seulement si elles perdaient le plasmide. Donc l'absence de croissance indiquait la dépendance de la levure pour le plasmide

contenant RoBP1. La levure dépendante de RoBP1 pouvait croître sur le milieu 5'FOA si elle était transformée avec un autre plasmide contenant RoBP1, ce qui confirmait la dépendance à RoBP1 et non au plasmide.

1.2.2 Tests de complémentation pour trouver l'homologue de RoBP1

Des tests de complémentation ont été effectués afin de déterminer quels gènes de levure la mutagenèse avait rendu inopérants et ainsi entraîné la dépendance à RoBP1. La méthode consistait à transformer une bibliothèque génomique de levure dans la souche dépendante de RoBP1 pour ensuite récupérer le ou les plasmides de la bibliothèque permettant la croissance sur le milieu 5'FOA. Ce milieu force le rejet du plasmide contenant RoBP1 et permet la croissance de la souche dépendante de RoBP1 seulement si le plasmide transformé contient un fragment d'ADN pouvant compléter pour la perte de RoBP1, donc contenant un gène pouvant remplacer la fonction de RoBP1. Deux régions des chromosomes IV et XI complètent pour la perte du plasmide codant pour RoBP1 et les gènes présents ont été identifiés par séquençage (Tableau 1). Sur le chromosome XI, le gène CAF4 a été identifié plus précisément par complémentation.

La stratégie utilisée nécessite que la mutation de l'homologue fonctionnel de RoBP1 soit létale pour la levure. La létalité observée ici semble provenir de mutations dans 2 gènes différents, comme le suggèrent les expériences préliminaires de complémentation effectuées par Jim Boulanger. Ces 2 mutations seraient viables lorsque présentes dans 2 souches distinctes, mais la combinaison des 2 mutations dans

la même souche serait létale. Ce concept est nommé létalité synthétique. Les homologues fonctionnels pourraient donc être deux protéines aux activités redondantes ou faisant partie de deux voies métaboliques redondantes mais convergentes.

1.3 Les pseudogènes hY

1.3.1 L'abondance de séquences génomiques homologues aux ARNs Y

Les gènes codant pour les différents ARNs Y chez l'humain et chez d'autres organismes moins complexes comme la souris semblent être présents en une seule copie par génome (WOLIN & STEITZ, 1983; MARAIA et al., 1994; MARAIA et al., 1996; FARRIS et al., 1996). Par contre, l'existence de plusieurs séquences génomiques homologues aux ARNs Y est connue depuis longtemps. Ces séquences ont notamment compliqué le clonage des gènes fonctionnels. Des expériences de buvardage de type Southern montrent une abondance de séquences génomiques homologues aux ARNs Y chez l'humain et, dans une moindre mesure, chez d'autres organismes (O'BRIAN & HARLEY, 1992; FARRIS et al., 1995).

Seulement quelques-unes de ces séquences furent caractérisées. Une séquence homologue à l'ARN hY3 fut identifiée dans le gène de l'alpha-globine 1 (JURKA et al., 1988). D'autres loci contiennent des séquences homologues à l'ARN hY4 (O'BRIAN & HARLEY, 1992). Des séquences homologues à l'ARN Y1 furent

Chromosome	Gène	Description
IV	UBC5	enzyme de conjugaison de l'ubiquitine
	PST1	organisation et biogénèse de la paroi cellulaire
	TGL2	lipase de triglycéride
	YDR057W	?
	YDR056C	?
	MAK21 (60%)	assemblage de la sous-unité ribosomale 60S
XI	SPC34	constituant du cytosquelette
	CAF4	composant du complexe CCR4-NOT, régulation de la transcription
	DID2	transport vacuolaire
	YKR035C	?
	YKR038C	?

Tableau 1 : Gènes présents sur les fragments d'ADN complétant la perte de RoBP1.

également caractérisées chez la souris (CROUCH & LIEBKE, 1989) et le rat (PRUIJN et al., 1993). Lorsque comparées aux séquences des gènes fonctionnels, les séquences de ces pseudogènes présentent quelques nucléotides différents et ne présentent aucun promoteur typique apparent. La région d'homologie se limite à la simple séquence de l'ARN Y. Donc, ces séquences, également nommées pseudogènes, ne sont vraisemblablement pas transcrites de façon autonome.

Cette abondance de pseudogènes n'est pas un phénomène unique aux ARNs hY. L'existence de telles séquences homologues, comme par exemple pour les petits ARNs nucléaires U1-6, est également observée depuis longtemps dans le génome humain et dans d'autres génomes de mammifères (DENISON et al., 1981). En fait, la majorité des ARNs cellulaires ou nucléaires transcrits par la pol I, II ou III (par exemple les gènes pour les ARNt, ARNr 5S, ARN 7SL et ARN 7SK), présentent de telles séquences génomiques à des niveaux d'abondance variables (WEINER et al., 1986). Notons que contrairement aux ARNs Y, quelques-unes de ces séquences sont des copies fonctionnelles du même gène. Une meilleure estimation du nombre réel des pseudogènes a été rendue possible grâce au séquençage du génome humain (LANDER et al., 2001).

1.3.2 Les éléments transposables dans le génome

La présence de séquences répétées dans les génomes eucaryotes est connue depuis longtemps (BRITTEN & KOHNE, 1968). Aujourd'hui il est admis que plus de 50% du génome humain est composé de séquences répétées, tandis que les

séquences codantes comptent pour environ 1%. De façon générale, ces répétitions peuvent se diviser en 3 grands groupes : (1) les répétitions dérivées d'éléments transposables; (2) les répétitions en tandem de séquences simples, aussi appelées microsatellites, minisatellites ou satellites, dont la taille et le nombre sont très variables; et (3) les duplications de segments, qui consistent en des séquences de 1-300 kb copiées d'une région du génome à une autre (LANDER et al., 2001; JASINSKA & KRZYZOSIAK, 2004).

Parmi toutes ces séquences répétées, les éléments transposables (ou éléments mobiles) et leurs vestiges sont les plus représentés avec près de 45% du génome humain. Ces éléments transposables sont décrits ci-dessous et leurs caractéristiques principales sont résumées dans le tableau 2. Les 2 grandes classes d'éléments mobiles sont les rétrotransposons, qui s'intègrent à de nouvelles locations génomiques via un intermédiaire ARN, et les transposons à ADN.

Les transposons à ADN sont similaires aux transposons bactériens et occupent environ 3% du génome. Ces éléments s'excisent d'un site génomique et s'intègrent dans un autre site ailleurs dans le génome par un mécanisme « couper et coller » (LANDER et al., 2001). Les transposons à ADN ne sont donc pas dupliqués. Ces éléments sont flanqués par des inversions terminales répétées (ITR) et de courtes répétitions directes. Les transposons à ADN codent pour une transposase qui est retrouvée liée aux ITRs et est responsable de la mobilité de ces éléments. La spécificité de la séquence d'intégration étant limitée à quelques nucléotides, l'insertion peut se faire à plusieurs endroits du génome, mais elle se fait habituellement à proximité (<100 kb) de l'insertion parentale (KAZAZIAN, 2004).

Type d'élément transposable			Caractéristiques
Transposons à ADN			<ul style="list-style-type: none"> ▶ 3% du génome ▶ mécanisme « couper et coller » ▶ flanqués d'inversions terminales répétées et de répétitions directes ▶ codent pour une transposase
Rétrotransposons	Autonomes	LTR	<ul style="list-style-type: none"> ▶ 8% du génome ▶ mécanisme « copier et coller » ▶ flanqués de longues répétitions terminales ▶ codent pour une transcriptase inverse et endonucléase ▶ ex : Rétrovirus endogènes
		Non-LTR (LINES)	<ul style="list-style-type: none"> ▶ 21% du génome ▶ mécanisme « copier et coller » ▶ flanqués de duplications du site d'insertion ▶ taille de 6-8 kb ▶ codent pour une transcriptase inverse et endonucléase ▶ ex : L1
	Non-autonomes (SINES)		<ul style="list-style-type: none"> ▶ 13% du génome ▶ mécanisme « copier et coller » ▶ flanqués de duplications du site d'insertion ▶ taille de 90-300 pb ▶ ne codent pour aucune protéine ▶ ex : Alu

Tableau 2 : Caractéristiques des différents types d'éléments transposables dans le génome humain.

La plupart de ces éléments seraient aujourd'hui inactifs et les séquences retrouvées seraient des vestiges d'anciens éléments.

Un rétrotransposon est d'abord transcrit en ARN, rétrotranscrit en ADN, et finalement réintégré ailleurs dans le génome par un mécanisme « copier et coller ». L'élément est ainsi dupliqué. Il existe 2 classes majeures de rétrotransposons autonomes dans les génomes eucaryotes, et plusieurs types de rétrotransposons non-autonomes qui dépendent des éléments autonomes pour leur rétrotransposition.

Les rétrotransposons autonomes à longues répétitions terminales (LTR) et leurs vestiges comptent pour environ 8% du génome humain. Parmi la variété de rétrotransposons LTR, notons les différentes classes de rétrovirus endogènes (ERVs) (LANDER et al., 2001). Leur structure est en effet très similaire à celle des rétrovirus. Ils codent pour plusieurs protéines fournissant les activités nécessaires à la rétrotransposition et tous les éléments de régulation de leur transcription sont contenus au niveau des LTRs. La transcription inverse des rétrotransposons LTR est un processus complexe qui a lieu au cytoplasme dans une particule ressemblant à un virus et utilise une amorce d'ARNt (WHITCOMB & HUGHES, 1992). Mais, contrairement aux rétrovirus, les rétrotransposons LTR ne produisent pas de protéine d'enveloppe fonctionnelle.

L'autre classe majeure de rétrotransposons autonomes ne contient pas de LTRs. Ces éléments sont connus sous le nom de LINEs (long interspersed nucleotide elements) et occupent près de 21% du génome humain. Ces rétrotransposons, d'une taille variant entre 6 et 8 kb, sont très anciens et des homologues sont retrouvés dans le génome d'une grande variété d'organismes eucaryotes (SINGER, 1982;

EICKBUSH, 1992; LANDER et al., 2001). Les LINEs sont transcrits par la pol II et polyadénylés, codent pour une transcriptase inverse nécessaire à leur réplication et sont flanqués de courtes duplications du site d'insertion (TSDs). La transcription inverse des LINEs se fait directement dans le noyau au niveau de l'ADN génomique par un processus nommé « target primed reverse transcription » (TPRT) qui sera décrit plus loin. La plupart des éléments retrouvés sont largement tronqués à leur extrémité 5' et sont donc inactifs. Dans le génome humain il existe 3 familles de LINEs apparentés : LINE1, LINE2 et LINE3. Seulement LINE1, aussi appelé L1, serait encore actif aujourd'hui (LANDER et al., 2001).

La plus importante classe d'éléments mobiles non-autonomes est formée par les SINEs (short interspersed nucleotide elements), qui occupent environ 13% du génome humain (SINGER, 1982; LANDER et al., 2001). Ces éléments de 90-300 pb ne codant pour aucune protéine, leur rétrotransposition nécessite l'action de protéines encodées par d'autres rétrotransposons autonomes. De façon générale, les SINEs sont transcrits par la pol III via un promoteur interne et, tout comme les LINEs, montrent une queue polyadénylée et sont flanqués par des TSDs. La plupart des SINEs seraient dérivés de différents gènes transcrits par la pol III comme les ARNt et l'ARN 7SL (WEINER et al., 1986). Les SINEs sont ubiquitaires chez les mammifères, mais peuvent également se retrouver chez d'autres organismes. Parmi la variété de SINEs, notons l'élément MIR, les éléments Alu et SVA chez l'humain, et les éléments B1 et B2 chez les rongeurs (LANDER et al., 2001; WATERSTON et al., 2002; OSTERTAG et al., 2003).

1.3.3 L'élément L1

L'élément L1 (LINE1) est le rétrotransposon le plus important des LINES. Avec ses 520 000 copies, il compte pour environ 17% du génome humain (LANDER et al., 2001). La vaste majorité (plus de 99,8%) des séquences L1 ne sont pas mobiles, mais les génomes humains et murins contiennent environ 60-100 et 3000 L1 actifs, respectivement (GOODIER et al., 2001; BROUHA et al., 2003). Un élément L1 complet et fonctionnel est d'une taille de 6kb et contient un promoteur interne situé dans la région 5'UTR, un court 3'UTR, un signal de polyadénylation suivi immédiatement d'une queue poly(A), et l'élément entier est flanqué de TSDs (Figure 2; OSTERTAG & KAZAZIAN, 2001). L'élément L1 code également pour 2 protéines, ORF1 et ORF2, nécessaires à la rétrotransposition et formant un complexe ribonucléoprotéique avec l'ARN L1. La protéine ORF2 présente 2 domaines ayant des activités de transcriptase inverse et d'endonucléase (MATHIAS et al., 1991; FENG et al., 1996; MORAN et al., 1996). La grande majorité des éléments L1 retrouvés dans le génome sont incomplets et tronqués à l'extrémité 5'. L'abondance de ces éléments L1 non-fonctionnels pourrait s'expliquer partiellement par une faible processivité de la protéine ORF2 qui atteint rarement l'extrémité 5' lors de la transcription inverse, ou par l'activité de la RNase H cellulaire (OSTERTAG & KAZAZIAN, 2001). ORF1 est une protéine pouvant lier l'ARN et possède une activité de chaperonne pour les acides nucléiques, mais sa fonction dans la rétrotransposition de L1 est inconnue (MARTIN & BUSHMAN, 2001; KOLOSHA & MARTIN, 2003).

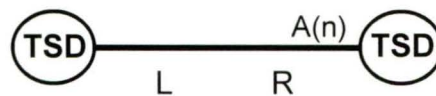


Figure 2 : Structure des rétrotransposons L1 et Alu.

L'élément L1 est flanqué de duplications du site d'insertion (TSD), possède des régions non-traduites en 5' (5'UTR) et 3' (3'UTR), un signal de polyadénylation (AATAAAA) et une queue polyadénylée (A(n)). L1 code pour 2 protéines, ORF1 et ORF2. La protéine ORF2 présente une activité de transcriptase inverse (RT) et d'endonucléase (EN). L'élément Alu est formé de 2 monomères similaires (L et R), possède une queue polyadénylée et est flanqué de duplications du site d'insertion. Adapté de OSTERTAG & KAZAZIAN, 2001.

Les éléments L1 sont retrouvés principalement dans des régions génomiques riches en AT et pauvres en gènes. Cette distribution pourrait s'expliquer par une sélection négative contre les insertions d'éléments L1 dans des gènes (KORENBERG & RYKOWSKI, 1988; LANDER et al., 2001).

1.3.4 L'élément Alu

Le membre le plus important de la classe des SINEs est l'élément Alu qui se retrouve uniquement chez les primates. Cet élément, connu depuis plus de 25 ans, tire son nom du site de restriction enzymatique *AluI* qu'il contient (HOUCK et al., 1979). Ses 1,1 millions de copies comptent pour près de 11% du génome humain et en font l'élément mobile le plus abondant (LANDER et al., 2001). L'élément Alu consiste en une séquence d'environ 300 pb formée de 2 monomères similaires séparés par une courte série d'adénines (Figure 2). L'extrémité 3' de l'élément Alu se termine avec une séquence riche en adénines, la taille totale de l'élément Alu dépendant de la taille variable de cette extrémité, habituellement 20 à 100 nucléotides (ROY-ENGEL et al., 2002). Comme mentionné précédemment pour les SINEs, l'élément Alu est flanqué par de courtes répétitions directes. Les 2 monomères sont indépendamment dérivés de l'ARN 7SL et l'élément Alu serait issu de la fusion de ces 2 monomères il y a plus de 60 millions d'années, soit environ au moment de la radiation des primates (QUENTIN, 1992). Les éléments Alu se divisent en sous-familles différant au niveau de leur séquence et dans l'âge de leur apparition durant l'évolution des primates.

Seulement certains éléments Alu, issus de la sous-famille la plus récente, seraient encore actifs aujourd'hui (BATZER & DEININGER, 2002).

Les séquences Alu ne sont pas distribuées uniformément dans le génome humain puisqu'elles s'accumulent préférentiellement dans les régions riches en gènes et donc riches en GC (KORENBERG & RYKOWSKI, 1988; LANDER et al., 2001; GROVER et al., 2004). Comme les éléments Alu les plus récents et les éléments L1 ne montrent pas ce biais pour les régions riches en GC, il est peu probable qu'une préférence dans le lieu d'insertion soit responsable de cette distribution génomique. Elle pourrait être expliquée par une sélection positive en faveur d'une fonction apportée par les éléments Alu situés dans les régions riches en gènes (LANDER et al., 2001). Mais il est aussi probable que le mécanisme d'élimination de ces éléments dans des régions riches en gènes confère un désavantage sélectif, résultant en un taux plus élevé de perte de ces éléments dans les régions pauvres en gènes (BROOKFIELD, 2001; BATZER & DEININGER, 2002).

1.3.5 Le mécanisme de rétrotransposition des LINEs et des SINEs

Les caractéristiques structurales des rétrotransposons mènent à la conclusion que l'information sous forme d'ARN est insérée dans des bris chromosomiques espacés (staggered chromosomal breaks; WEINER et al., 1986). Par contre, le mécanisme détaillé de rétrotransposition, impliquant une activité enzymatique, a été partiellement clarifié par différentes études ultérieures (LUAN et al., 1993; JURKA, 1997; FENG et al., 1996; COST et al., 2002). Ainsi, la rétrotransposition de L1 (et

des LINEs) serait effectuée par un processus de transcription inverse et d'intégration couplé nommé TPRT (Figure 3). Brièvement, l'activité endonucléase de la protéine ORF2 coupe un brin d'ADN au site cible, produisant un brin d'ADN libre avec une extrémité 3'OH. La séquence du site cible reconnue par la protéine est l'hexanucléotide 5'-TTAAAA-3' et la coupure se fait au niveau du brin antisens, soit entre les nucléotides 3'-AA et TTTT-5'. Ensuite, cette extrémité TTTT du brin d'ADN coupé s'apparie à l'extrémité 3' polyadénylée de l'ARN L1. L'activité de transcriptase inverse de la protéine ORF2 utilise alors cet ARN comme modèle et l'extrémité 3'OH de l'ADN comme amorce pour la transcription inverse. Ensuite, le second brin d'ADN est coupé à environ une quinzaine de nucléotides en aval de la coupure initiale, l'ADNc est intégré, l'ARN est rejeté, et la synthèse de l'ADN est complétée. Le résultat est une copie du rétrotransposon initial intégrée à une nouvelle location génomique et flanquée par 2 courtes duplications du site d'insertion d'environ 15 nucléotides (OSTERTAG & KAZAZIAN, 2001).

Les SINEs utiliseraient la machinerie de rétrotransposition de L1 pour assurer leur mobilisation. En effet, bien que les protéines de l'élément L1 montrent une forte préférence pour mobiliser l'ARN L1 qui les encodent, elles peuvent également, à une fréquence beaucoup moindre, agir en *trans* pour promouvoir la rétrotransposition d'autres ARNs cellulaires (WEI et al., 2001). Cette mobilisation en *trans* a été démontrée pour d'autres rétrotransposons non-autonomes comme les éléments Alu (DEWANNIEUX et al., 2003) et pour un SINE présent chez l'anguille (KAJIKAWA & OKADA, 2002).

De plus, la machinerie de rétrotransposition de L1 est aussi responsable de la création de pseudogènes maturés (processed pseudogenes; ESNAULT et al., 2000). Ces pseudogènes sont générés lorsqu'un ARN messager est rétrotranscrit et intégré à une nouvelle position dans le génome par la machinerie de rétrotransposition de L1. On estime à environ 8000 le nombre total de ces pseudogènes maturés dans le génome humain (ZHANG et al., 2003) et les pseudogènes de protéines ribosomales seraient les plus abondants avec plus de 2000 copies (ZHANG et al., 2002).

La machinerie de rétrotransposition de L1 peut aussi participer à l'intégration de pseudogènes chimériques dans le génome. Cette classe de pseudogènes est générée lorsque, durant la transcription inverse, la protéine ORF2 passe de l'ARN L1 à un autre ARN cellulaire à proximité, pour ensuite poursuivre la transcription inverse. Il en résulte, après intégration, une fusion entre un élément L1 (ou Alu) et un autre ARN, principalement le petit ARN nucléaire U6 (BUZDIN et al., 2003).

1.3.6 L'influence des éléments mobiles

Depuis leur apparition jusqu'à aujourd'hui, les éléments mobiles ont affecté le génome de différentes façons, certaines bénéfiques et d'autres néfastes. L'analyse de la divergence des séquences des éléments transposables par rapport aux séquences consensus permet d'observer que globalement, l'activité des transposons a décliné considérablement depuis les 35-50 derniers millions d'années (LANDER et al., 2001). Aujourd'hui, on estime qu'un événement de rétrotransposition survient chez 1

individu sur 10-250 pour l'élément L1 (OSTERTAG & KAZAZIAN, 2001) et 1 sur 100-200 pour l'élément Alu (DEININGER & BATZER, 1999).

La rétrotransposition des éléments L1, Alu, SVA ou autres peut entraîner une instabilité génétique et provoquer des maladies génétiques (KAZAZIAN et al., 1988; DEININGER & BATZER, 1999; SYMER et al., 2002; OSTERTAG et al., 2003). En effet, la rétrotransposition d'un élément mobile dans un gène peut avoir différents effets, dépendamment du lieu d'insertion. Une insertion dans la portion 5' non-traduite (5'UTR) peut altérer l'expression du gène, tandis qu'une insertion dans un exon peut modifier le cadre de lecture et entraîner la production d'une protéine non-fonctionnelle. Un élément mobile peut également altérer l'épissage de l'ARN messager s'il s'insère à proximité d'un site d'épissage. Par contre, une insertion d'un élément mobile dans un intron peut survenir sans qu'il y ait d'effet immédiat sur le gène. De plus, une recombinaison homologue inégale entre des séquences Alu ou L1 présentes dans des gènes peut entraîner des délétions et des duplications de séquences qui sont néfastes (DEININGER & BATZER, 1999; OSTERTAG & KAZAZIAN, 2001). Ainsi, les insertions d'éléments Alu comptent pour environ 0,3% de toutes les maladies génétiques humaines.

Les rétrotransposons ont également contribué à l'expansion, à la diversification et à l'évolution des génomes. En plus de la séquence du rétrotransposon elle-même, L1 peut mobiliser un segment d'ADN supplémentaire adjacent en 3', dû à un signal de polyadénylation et de clivage peu efficace. Ce phénomène de « transduction » de séquences surviendrait dans 10-20% des événements de rétrotransposition d'éléments L1. Ces événements ont le potentiel

d'introduire des exons et des séquences régulatrices à de nouvelles positions génomiques et ainsi de créer de nouveaux gènes (MORAN et al., 1999; GOODIER et al., 2000; PICKERAL et al., 2000). De plus, il a été montré récemment que plus de 5% des exons internes épissés de façon alternative sont dérivés d'éléments Alu. Ces exons dérivés d'éléments Alu auraient été créés exclusivement via l'exonisation d'éléments introniques (SOREK et al., 2002; LEV-MAOR et al., 2003).

Finalement, l'insertion d'un élément L1 au niveau d'une cassure d'ADN double-brin pourrait réparer la lésion, indépendamment d'une activité endonucléase. L'extrémité 3'OH d'un brin d'ADN libéré au niveau de la cassure pourrait servir d'amorce pour la transcription inverse de l'ARN L1 (MORRISH et al., 2002).

Lors du clonage du gène fonctionnel de l'ARN hY5, l'analyse des séquences flanquantes a révélé des signatures de rétrotransposition et il a été proposé que ce gène soit dérivé d'un événement de rétrotransposition de l'ARN hY3 (MARAIA et al., 1996). Plusieurs pseudogènes de l'ARN hY4 et de l'ARN hY3 ont été retrouvés associés à des éléments Alu dans le génome (JURKA et al., 1988; O'BRIAN & HARLEY, 1992), ce qui permet de supposer l'existence de pseudogènes chimériques Alu-hY. Afin de mieux comprendre leur origine et leur fonction, il s'avèrait donc nécessaire d'effectuer le décompte précis et d'analyser rigoureusement les séquences de ces pseudogènes des ARN hY.

1.4. Les RNPs Ro et l'épissage

1.4.1 L'épissage constitutif et alternatif

L'épissage est un processus nucléaire qui consiste en l'excision des séquences non-codantes introniques du transcrit primaire (ARN pré-messager) et la réunion bout à bout des exons. Le résultat final de ce processus est un ARN messager prêt à être exporté vers le cytoplasme pour y être traduit en protéine. L'épissage est un ensemble de réactions catalysées par un large complexe ribonucléoprotéique: le spliceosome. Ce dernier est composé de 5 petits ARNs nucléaires (snRNA) riches en uracile : les ARNs U1, U2, U4, U5 et U6, chacun fonctionnant en complexe avec des protéines pour former de petites ribonucléoprotéines nucléaires (snRNPs). Une multitude d'autres protéines viennent s'ajouter aux snRNPs pour former le spliceosome. Des études avancent même le nombre de 300 protéines qui feraient partie du spliceosome (JURICA et al., 2003).

Le spliceosome reconnaît les jonctions entre les exons et les introns (les sites d'épissage 5' et 3') au niveau de l'ARN pré-messager. La jonction exon/intron du site 5' est variable, mais possède généralement la séquence consensus AG/GURAGU et la séquence consensus de la jonction intron/exon du site 3' est YAG/R, où les nucléotides en gras sont les mieux conservés (MOUNT, 1982; PADGETT et al., 1986; SHAPIRO & SENAPATHY, 1987). Un autre élément aidant à la définition du site d'épissage 3' est le site de branchement, qui est habituellement suivi d'une région riche en pyrimidines. Le site de branchement, dont la séquence consensus est

YURAY, est habituellement situé à 10-50 nucléotides en amont du site d'épissage 3' (HARRIS & SENAPATHY, 1990).

L'épissage est un processus très complexe dont il reste encore aujourd'hui à élucider les mécanismes détaillés. Initialement, le site d'épissage 5' est reconnu par le snRNP U1 par interaction ARN-ARN. La région de branchement et le site d'épissage 3' sont définis par association avec la protéine mBBP/SF1, l'hétérodimère U2AF65/35 et certaines protéines SR, pour former le complexe d'engagement (BURGE et al., 1999). L'interaction entre les facteurs reconnaissant le site 3' et ceux reconnaissant le site 5' permet la définition de l'exon (ROBBERTSON et al., 1990). La région 3' est ensuite reconnue par interaction ARN-ARN avec le snRNP U2. La formation du spliceosome mature est accomplie ensuite par le recrutement du tri-snRNP U4/U6/U5. L'épissage de l'ARN pré-messager en ARN messenger mature est obtenu après plusieurs réarrangements des snRNPs et 2 étapes séquentielles de trans-estérification (BURGE et al., 1999). Dans la première, le 2'OH du nucléotide A présent au site de branchement attaque le phosphate au site d'épissage 5', libérant l'extrémité de l'exon et un intermédiaire en forme de lasso. Dans la deuxième étape, le 3'OH de l'exon libre attaque le phosphate au site d'épissage 3', résultant en un ARN messenger ligué et un intron relâché sous forme de lasso (ALBERTS et al., 1994).

Les récents progrès dans le domaine de la génomique ont montré que le nombre de gènes n'est pas nécessairement proportionnel à la complexité d'un organisme et à la taille de son protéome. L'épissage alternatif, par opposition à l'épissage constitutif, peut expliquer la grande diversité de protéines différentes pouvant être produites à partir d'un nombre limité de gènes. Un exemple frappant est

le gène DSCAM chez *Drosophila melanogaster* qui a la capacité de produire plus de 38 000 isoformes de la protéine par épissage alternatif, soit plus de 2 fois le nombre total de gènes de l'organisme (SCHMUCKER et al., 2000). Selon les premières estimations, des évidences d'épissage alternatif étaient observées dans 35% des gènes (MIRONOV et al., 1999). Mais des études récentes suggèrent que plus de 74% des gènes sont épissés de façon alternative (JOHNSON et al., 2005).

L'utilisation alternative des sites d'épissage 5' ou 3' permet d'obtenir, à partir d'une seule séquence d'ARN pré-messager, plusieurs combinaisons de séquences d'ARN messenger. De cette façon, un exon donné peut être entièrement inclu ou exclu d'un ARN messenger. Des exons peuvent également avoir plusieurs sites d'épissage 5' ou 3' utilisés de façon alternative, ce qui permet d'allonger ou de raccourcir les séquences exoniques. D'autres exons peuvent être mutuellement exclusifs de sorte qu'un seul d'entre eux peut être présent dans l'ARN messenger mature. L'épissage alternatif peut aussi faire en sorte qu'une partie ou la totalité d'une séquence intronique sera retenue dans l'ARN messenger mature et éventuellement traduite en protéine (GRAVELEY, 2001; BLACK, 2003). Dans certains cas, l'épissage alternatif peut même mener à l'introduction d'un codon de terminaison prématuré, dirigeant l'ARN vers la dégradation par « nonsense-mediated decay » (NMD) plutôt que vers la traduction (LEWIS et al., 2003).

Les protéines ainsi générées, ayant des séquences peptidiques différentes, ont vraisemblablement des structures et des fonctions différentes. Les isoformes peuvent différer au niveau de l'activité enzymatique, de la localisation cellulaire ou de l'affinité pour un ARN ou une autre protéine. Deux isoformes d'une même protéine

peuvent même avoir des activités antagonistes. Par exemple, l'épissage alternatif du gène Bcl-x, un membre de la famille du gène Bcl2, produit plusieurs isoformes dont une isoforme longue (Bcl-xL), possédant une activité anti-apoptotique, et une isoforme courte (Bcl-xS), possédant une activité pro-apoptotique (BOISE et al., 1993).

1.4.2 La régulation de l'épissage alternatif

La possibilité de produire plusieurs ARNm pouvant conduire, après traduction, à des protéines ayant des activités distinctes à partir d'un même gène nécessite un contrôle très serré. De plus, ce contrôle fait en sorte que certains événements d'épissage alternatif se produisent préférentiellement dans des types cellulaires particuliers, ou à un moment précis du développement ou sous des conditions physiologiques spécifiques (GRAVELEY, 2001). Généralement, la régulation de l'épissage alternatif se fait au niveau de la reconnaissance des sites d'épissage. Les événements d'épissage alternatif impliquent souvent une compétition entre plusieurs sites d'épissage faibles dont les séquences divergent plus ou moins des séquences consensus. Au niveau des exons et des introns, certains éléments de séquence spécifiques favorisent l'association de certains composants du spliceosome, tandis que d'autres éléments recrutent des facteurs nuisant à la reconnaissance des sites d'épissage. Globalement, il existe donc 4 classes d'éléments régulateurs de l'épissage : les éléments exoniques de régulation positive (*exonic splicing enhancers* : ESE) et négative (*exonic splicing silencers* : ESS) et les éléments introniques de

régulation positive (*intronic splicing enhancers* : ISE) et négative (*intronic splicing silencers* : ISS) (BLACK, 2003). Ces éléments sont reconnus par de nombreux facteurs d'épissage dont certains groupes de protéines comme les protéines contenant des domaines riches en sérine-arginine (protéines SR) ou les ribonucléoprotéines nucléaires hétérogènes (protéines hnRNP). Donc, l'activation ou la répression de chaque événement d'épissage résulte de la liaison de certains facteurs d'épissage aux séquences régulatrices présentes dans les ARN pré-messagers, combinée avec la force de chaque site d'épissage.

1.4.2.1 Les éléments régulateurs exoniques

Les séquences régulatrices situées dans les exons peuvent avoir un effet positif ou négatif sur la sélection de sites d'épissage alternatif. La plupart des éléments positifs (ESE) connus sont liés par des membres de la famille des protéines SR. Ces facteurs d'épissage très étudiés sont impliqués autant au niveau de l'épissage constitutif qu'au niveau de l'épissage alternatif. Les membres de cette famille contiennent au moins un motif RRM et un domaine riche en arginine/sérine (domaine SR). La liaison des protéines SR aux ESE se fait via le motif RRM et l'activation de l'épissage est réalisée par le recrutement de composants du spliceosome par interaction protéine-protéine via les domaines SR (BLENCOWE, 2000). De façon générale, les ESE peuvent favoriser la liaison de U2AF et du snRNP U2 aux sites d'épissage 3' faible, ou l'association du snRNP U1 aux sites d'épissage 5' faibles (COTE et al., 1999a; GRAVELEY, 2000; GRAVELEY et al., 2001). Dans certains

autres cas, la liaison du facteur d'épissage aux ESE permet l'activation de l'épissage en renversant l'effet d'un élément inhibiteur (KAN & GREEN, 1999).

Les éléments exoniques ayant une action inhibitrice sur l'épissage (ESS) sont reconnus pour la plupart par des protéines appartenant au vaste groupe des protéines hnRNP. Parmi les mieux étudiées, hnRNP A1 est une protéine abondante possédant 2 domaines de liaison à l'ARN du type RRM et un domaine riche en glycine. Cette protéine réprime l'épissage dans plusieurs ARN pré-messagers, dont FGFR2, Tat et c-Src (DEL GATTO-KONCZAK et al., 1999; ZHU et al., 2001; ROOKE et al., 2003) par association à des ESS ayant des séquences similaires à la séquence consensus UAGGGA/U (BURD & DREYFUSS, 1994). Les protéines très similaires hnRNP H et hnRNP F peuvent également agir comme inhibiteurs de l'épissage en reconnaissant des ESS riches en G (CHEN et al., 1999; JACQUENET et al., 2001; CAPUTI & ZAHLER, 2001). Dans plusieurs cas, l'action des protéines hnRNP liées aux ESS s'oppose à l'effet des protéines SR associées aux ESE adjacents et le patron d'épissage résultant dépend de l'équilibre entre ces deux éléments (EPERON et al., 2000).

1.4.2.2 Les éléments régulateurs introniques

Des séquences situées dans des introns peuvent également être impliquées dans la régulation positive ou négative de l'épissage alternatif (ISE ou ISS). Ces éléments régulateurs sont souvent situés dans la région riche en pyrimidines du site d'épissage 3' ou près du site de branchement ou du site d'épissage 5', mais peuvent

aussi se retrouver à plusieurs centaines de nucléotides de l'exon alternatif (LIM & SHARP, 1998; BLACK, 2003). Tout comme les ESE, certains ISE sont reconnus par des protéines SR et sont riches en purines, mais d'autres peuvent également être riches en pyrimidines (GALLEGO et al., 1997; HASTINGS et al., 2001). D'autres ISE, constitués de séquences riches en uridines, sont nécessaires à l'activation de sites d'épissage 5' faibles en recrutant, par exemple, la protéine TIA-1 qui favorise l'association avec le snRNP U1 (FORCH et al., 2000). Les protéines hnRNP F/H peuvent aussi activer l'épissage en association avec certains ISE (CHEN et al., 1999).

Une des protéines les mieux étudiée pour son rôle dans la régulation négative de l'épissage alternatif au niveau des introns est la protéine hnRNP A1. La liaison de hnRNP A1 à un ISS, ainsi qu'à un ESS adjacent, contribue à l'inhibition de l'épissage de l'exon 3 du gène Tat du VIH-1 en bloquant l'association du snRNP U2 avec le site de branchement (TANGE et al., 2001). La protéine hnRNP A1 est également impliquée dans la régulation de l'épissage alternatif de son propre ARN pré-messager. La présence de l'exon alternatif 7B dans l'ARNm donne l'isoforme hnRNP A1B, tandis que l'absence de cet exon produit la protéine hnRNP A1. La régulation de l'inclusion de cet exon nécessite la présence de plusieurs sites de liaison pour hnRNP A1 situés de part et d'autre de l'exon 7B. La liaison de hnRNP A1 à ces sites n'empêchant pas la liaison du snRNP U1 au site d'épissage 5', un mécanisme de « looping out » est plutôt proposé. Selon ce modèle, les molécules de hnRNP A1 flanquant l'exon 7B pourraient multimériser, rapprochant les sites d'épissage distants et excluant ainsi la région contenant l'exon (BLANCHETTE & CHABOT, 1999; NASIM et al., 2002). Finalement, les protéines SR ASF/SF2 et SRp30c sont aussi

responsables de la régulation négative de l'épissage lorsque recrutées au niveau d'éléments introniques particuliers (KANOPKA et al., 1996; SIMARD & CHABOT, 2002).

En conclusion, le patron d'épissage d'un ARN donné dépend de la position des nombreux éléments activateurs et inhibiteurs, de la nature et de la quantité des facteurs qui y sont associés, ainsi que de la force et de la position des sites d'épissage en compétition (HASTINGS et al., 2001).

1.4.3 Les protéines associées à la RNP Ro et l'épissage.

Aucune fonction claire n'étant encore attribuée à la particule complexe qu'est la RNP Ro, les recherches se concentrent sur l'étude des composantes protéiques isolées. Les protéines identifiées jusqu'à maintenant comme interagissant avec la RNP Ro démontrent cependant des rôles cellulaires très variés. Sachant que ces protéines peuvent se lier aux séquences correspondant aux ARN hY se retrouvant dans des introns à proximité de sites d'épissage, il est intéressant de noter que parmi ces protéines, on en remarque 3 qui seraient impliquées dans le processus de l'épissage : PTB, hnRNP K et RoBP1.

1.4.3.1 La protéine PTB

Le protéine PTB est reconnue comme une protéine importante pour la répression de l'épissage de plusieurs exons alternatifs. La liaison de PTB à des

séquences introniques riches en pyrimidines est impliquée dans la répression de l'épissage alternatif de plusieurs exons tissus-spécifiques. Ces sites de liaison sont souvent situés dans la région riche en pyrimidines du site d'épissage 3', mais peuvent aussi être présents en d'autres endroits sur l'ARN pré-messager. Il a été suggéré que la liaison de PTB à la région de polypyrimidine du site d'épissage 3' pourrait interférer directement avec la liaison de U2AF et ainsi inhiber la reconnaissance du site et l'épissage de l'exon (LIN & PATTON, 1995; SINGH et al., 1995). Un seul site de liaison pour PTB peut dans certains cas être nécessaire à la répression d'un exon (SHEN et al., 2004). Cependant, dans de nombreux transcrits, plusieurs sites de liaison pour PTB situés de part et d'autre de l'exon régulé sont nécessaires à la répression. Par exemple, l'exon IIIb du gène FGFR2 est réprimé dans les tissus mésenchymateux par les nombreux sites de liaison pour PTB situés des 2 côtés de cet exon (CARSTENS et al., 2000; WAGNER & GARCIA-BLANCO, 2002). De même, l'exon SM du gène de l' α -actinine peut être réprimé dans tous les tissus excepté les tissus musculaires lisses grâce à un réseau de sites de liaison pour PTB autour de l'exon (SOUTHBY et al., 1999), et la liaison de PTB aux sites flanquant l'exon N1 du gène c-src entraîne la répression de cet exon dans toutes les cellules autres que les cellules neuronales (CHAN & BLACK, 1997; CHOU et al., 2000).

Cependant, le mécanisme de répression par PTB reste en grande partie inconnu. Étant donné que la plupart des exons réprimés par PTB sont flanqués par des sites de liaison situés dans les 2 introns adjacents et que PTB peut multimériser, il a été proposé que les protéines PTB pourraient interagir d'un côté à l'autre de l'exon pour créer une zone de silence (WAGNER & GARCIA-BLANCO, 2001), ce qui

s'apparente au mécanisme de « looping-out » proposé pour hnRNP A1. Par contre, l'action de PTB est nécessaire mais non suffisante à la répression de la plupart de ces exons. La présence d'autres éléments de séquence de régulation est souvent nécessaire pour obtenir une répression efficace, et il est donc difficile de déterminer le rôle précis de PTB. Cependant, une étude récente avance que l'insertion de séquences de liaison pour PTB pourrait être suffisante pour la répression d'un exon donné, sans avoir l'aide de sites de liaison pour d'autres protéines régulatrices et sans interférer directement avec la liaison de U2AF (AMIR-AHMADY, 2005).

1.4.3.2 La protéine hnRNP K

La protéine hnRNP K serait impliquée dans l'épissage alternatif du gène de la β -tropomyosine du poulet. L'ARN pré-messager transcrit à partir de ce gène contient 2 exons mutuellement exclusifs : les exons 6A et 6B. Un élément de séquence riche en pyrimidines situé dans l'intron à 37 nucléotides en aval de l'exon 6A est nécessaire à l'activation de l'épissage de l'exon. Cette séquence régulatrice active l'épissage en recrutant le snRNP U1 au site d'épissage 5' de l'exon 6A. La liaison de la protéine hnRNP K à cet élément est nécessaire à la reconnaissance de l'exon par un complexe d'activation comprenant d'autres protéines, dont PTB (EXPERT-BEZANÇON et al., 2002). En plus de sa participation dans l'épissage alternatif du gène de la β -tropomyosine, hnRNP K pourrait tenir d'autres fonctions dans l'épissage par son interaction avec différents facteurs d'épissage comme 98G, SRp20 et hnRNP L

(SHNYREVA et al., 2000). Les évidences concernant le rôle de hnRNP K dans l'épissage restent cependant peu nombreuses.

1.4.3.3 La protéine RoBP1

La troisième protéine associée aux RNPs Ro et ayant vraisemblablement un rôle dans l'épissage est la protéine RoBP1. L'isoforme PUF60 (poly-U binding factor-60 kDa) est identifié comme étant un facteur d'épissage général. In vitro, cette protéine favorise l'épissage de plusieurs introns, en collaboration avec le facteur d'épissage U2AF65 et la protéine SR p54, en facilitant l'association du snRNP U2 avec l'ARN pré-messager (PAGE-MCCAW et al., 1999).

D'autres études s'attardent sur l'orthologue de PUF60/FIR/RoBP1 chez la drosophile : Hfp. Cette protéine a été décrite comme un facteur d'épissage impliqué dans la régulation de l'oogenèse. Contrairement à PUF60 qui semble avoir une influence plus générale, Hfp participerait à l'épissage constitutif et alternatif de seulement quelques uns des ARN messagers exprimés dans l'ovaire. L'homologue de Hfp chez le rat peut lier une courte séquence d'ARN régulatrice nécessaire pour l'épissage alternatif adéquat du gène codant pour le précurseur de la protéine amyloïde (APP). La protéine pourrait même interagir directement avec le facteur d'épissage U2AF65 (POLEEV et al., 2000).

2. HYPOTHÈSES ET BUTS

Levure et RoBP1

Les particularités biochimiques et immunologiques des RNPs Ro contenant la protéine RoBP1 suggèrent une fonction particulière. L'identification et l'étude de l'homologue fonctionnel de RoBP1 chez la levure *Saccharomyces cerevisiae* pourrait permettre la découverte de la fonction des RNPs Ro humaines.

Pseudogènes hY

Il existe une abondance de séquences homologues aux ARNs hY répétées dans le génome. La machinerie de rétrotransposition de l'élément L1 pourrait avoir été utilisée pour la mobilisation de ces séquences. Cette rétrotransposition laisse des caractéristiques identifiables dans le génome. L'analyse bioinformatique approfondie des séquences hY génomiques pourrait permettre de découvrir leur origine. La localisation génomique de ces séquences pourrait aussi pointer vers une fonction possible.

hY et épissage

Des protéines interagissant avec les RNPs Ro seraient également impliquées dans l'épissage (PTB, hnRNP K et RoBP1). Des séquences homologues aux ARN hY se retrouvant dans des introns près de sites d'épissage pourraient recruter ces protéines et ainsi moduler l'épissage. Nous avons donc introduit des séquences correspondant aux ARNs hY dans les introns de modèles d'épissage alternatif (hnRNP A1 et Bcl-x) pour vérifier leur impact sur l'épissage.

3. MATÉRIEL ET MÉTHODES

3.1. Manipulations de base de l'ADN

3.1.1 Culture bactérienne et transformation de plasmides

Les bactéries de la souche *Escherichia coli* DH5 α sont rendues compétentes par une méthode utilisant le chlorure de calcium (AUSUBEL et al., 1993), puis congelées à -80°C jusqu'à leur utilisation éventuelle. La transformation des plasmides en bactéries est réalisée par choc thermique et se déroule de la façon suivante. Premièrement, 25 à 100 μ l de bactéries compétentes sont incubées sur glace en présence du plasmide d'intérêt pendant 20 min. Elles sont ensuite incubées pendant 90 sec. à 42°C puis laissées sur glace pendant 90 sec. Un volume de 500 μ l de milieu LB (Luria-Bertani) stérilisé à l'autoclave est ajouté et le tout est incubé pendant 30 min. à 37°C. Le milieu de culture est ensuite retiré par centrifugation et le culot bactérien est resuspendu dans 100 μ l de milieu LB. Les bactéries sont ensuite étalées dans des boîtes de pétris contenant du milieu autoclavé LB auquel on a ajouté 1,5% d'agar et 50 μ g/ml d'ampicilline. Les pétris sont ensuite incubés durant une nuit (environ 16 heures) à 37°C pour permettre la croissance des bactéries transformées.

3.1.2 Purification de plasmides

Les colonies contenant le plasmide d'intérêt sont repiquées et ensemencées

dans un milieu LB liquide autoclavé contenant de l'ampicilline. La culture bactérienne croît pendant une nuit à 37°C sous agitation à 225 tours/min., puis les plasmides sont isolés par la méthode de lyse alcaline (SAMBROOK et al., 1989) ou par une variante du protocole (LE GOUILL et al., 1994).

3.1.3 Visualisation d'ADN sur gel d'agarose

Tous les fragments d'ADN (plasmides, produits de PCR et produits de digestion enzymatique) sont visualisés par électrophorèse sur gel d'agarose 0,8-2% (SOUTHERN, 1979). La taille des fragments d'ADN est estimée en faisant migrer parallèlement un marqueur de poids moléculaire de 1kb ou 100 pb (New England Biolabs). La migration de l'ADN est effectuée à 100-120 V pendant 30-60 min dans un tampon TAE (40 mM Tris-acétate, 1 mM EDTA). Le gel d'agarose est ensuite incubé dans du tampon TAE contenant 1 mg/ml de bromure d'éthidium, puis il est soumis à un rayonnement ultraviolet de 254 nm pour visualiser les bandes d'ADN.

3.1.4 Isolement de bandes d'ADN, digestion enzymatique et ligation

La bande contenant l'ADN d'intérêt est excisée du gel d'agarose suite à son identification sous rayonnement ultraviolet. L'ADN est ensuite extrait du gel et purifié à l'aide de la trousse *QIAquick Gel Extraction Kit* (Qiagen Inc.) en suivant les recommandations du fabricant. Les conditions utilisées pour les digestions enzymatiques sont celles suggérées par les manufacturiers (Amersham Biosciences ou

New England Biolabs). Suite à une digestion enzymatique, le plasmide d'intérêt peut être traité à la phosphatase alcaline intestinale de veau (Amersham Biosciences) en suivant les recommandations du fabricant. La ligation d'un insert dans un plasmide a été réalisée en utilisant la trousse *Rapid DNA Ligation Kit* (Roche Diagnostics) ou la *T4 DNA Ligase* (New England Biolabs) selon les recommandations du fabricant.

3.1.5 Amplifications par PCR

Les réactions de PCR ont été effectuées dans un appareil Perkin-Elmer 9600 (Applied Biosystems). Chaque amplification par PCR a été réalisée, sauf avis contraire, dans un volume total de 50 µl en utilisant 25-500 ng d'ADN comme matrice, 500 nM des amorces 5' et 3' (Gibco BRL/Life technologies), 0.2 mM de dNTP (Amersham Biosciences), 2,5 unités de *Pwo DNA polymerase* (Roche Diagnostics) ou de *Taq DNA polymerase* (New England Biolabs) et le tampon fourni par le fabricant, selon les recommandations. La température d'élongation est de 72°C et la température d'appariement est de 5°C en dessous du T_m des amorces. Le temps d'élongation est de 1 min. par kilobase du produit à obtenir.

Levure et RoBP1

3.2 Souches de levure utilisées

La souche de levure à partir de laquelle sont construites les autres souches

utilisées dans ce projet est la souche *Saccharomyces cerevisiae* BY4705 (BRACHMANN et al., 1998) et est fournie gracieusement par le Dr Raymund Wellinger (Département de microbiologie, Université de Sherbrooke). La souche diploïde possède le génotype *MATa/MATa*, *ade2Δ::HISG/ade2Δ::HISG*, *his3-Δ200/his3-Δ200*, *leu2-Δ0/leu2-Δ0*, *lys2-Δ0/lys2-Δ0*, *met15-Δ0/met15-Δ0*, *trp1-Δ63/trp1-Δ63* et *ura3-Δ0/ura3-Δ0*. Cette souche est utilisée pour la formation des souches de délétion. La souche de levure dépendante de RoBP1 créée par Jim Boulanger (1^{ère} génération) fut construite à partir de la souche haploïde BY4705 *MATa*. Cette souche inclut le plasmide pCu4X6/RoBP1, contenant les marqueurs de sélection URA3 et ADE2, et permettant l'expression de la protéine RoBP1. La souche dépendante de RoBP1 utilisée dans ce projet est de 2^e génération, car elle est issue d'un rétrocroisement avec la souche haploïde BY4705 *MATa*. La souche *Saccharomyces cerevisiae* BY4704, utilisée comme contrôle dans certaines expériences, possède le génotype *MATa*, *ade2Δ::HISG*, *his3-Δ200*, *leu2-Δ0*, *lys2-Δ0*, *met15-Δ0* et *trp1-Δ63*.

3.3 Culture des levures

Les levures sont habituellement cultivées à une température de 30°C en phase solide ou liquide et les milieux utilisés peuvent être riches ou restrictifs (ROSE et al., 1990). Plus spécifiquement, le milieu de culture YEPD (yeast extract peptone dextrose) est un milieu riche composé de 10g/L de yeast extract (Difco), 20g/L de bacto-peptone (Fisher Scientific Ltd) et 2% de dextrose, à pH 5,8. Le dextrose,

préalablement filtré à l'aide d'un filtre de 0,22µm (Millipore), est ajouté au milieu après stérilisation à l'autoclave sous la forme d'une solution concentrée à 40%.

Le milieu synthétique minimal SD est constitué de 1,2g/L de la base azotée YNB (yeast nitrogen base) sans sulfate d'ammonium ni acides aminés (Difco), 5g/L de sulfate d'ammonium, 10 g/L d'acide succinique, 6g/L de NaOH et 2% de dextrose. Le milieu synthétique complet SC, aussi appelé Yc ou « Drop out », est obtenu en supplémentant le milieu minimal du mélange d'acides aminés approprié (SHERMAN et al., 1986). Les acides aminés suivants sont présents à la concentration de 0,05 g/L : aspartate, histidine, isoleucine, méthionine, phénylalanine, proline, sérine, tyrosine et valine, tandis que les acides aminés suivants sont présents à la concentration de 0,1 g/L : adénine, arginine, cystéine, leucine, lysine, thréonine, tryptophane et uracile. En omettant un ou plusieurs acides aminés dans la composition de ce mélange, on obtient un milieu restrictif. La croissance sur milieu restrictif est employée pour la sélection des levures selon leurs besoins nutritionnels particuliers. Par exemple, un milieu restrictif dépourvu de leucine, nommé Yc-Leu, ne permettra pas la croissance des souches de levures auxotrophes pour cet acide aminé.

La forme solide de ces différents milieux peut être obtenue en ajoutant 1,5-2% de bacto-agar (Difco) au milieu liquide avant stérilisation à l'autoclave.

3.4 Transfert de levures d'un Pétri à un autre (replica plating)

Dans certaines expériences où il est nécessaire de transférer plusieurs cultures d'une gélose à une autre pour tester des caractéristiques comme les besoins

nutritionnels ou le mating-type, la méthode du « replica plating » est utilisée. Cette méthode permet de reproduire le patron de croissance d'un Pétri à un autre. Un cylindre en acier au diamètre légèrement inférieur à celui d'une boîte de Pétri est recouvert d'une membrane de velours préalablement stérilisée à l'autoclave. Un anneau élastique maintient le velours sur le cylindre. La boîte de Pétri maîtresse, contenant les différentes cultures/colonies à transférer, est inversée sur le velours ainsi disposé. Pour que des levures adhèrent au velours, une légère pression est exercée sur le Pétri avant de le retirer. Le patron de croissance des levures laissé sur le velours est transféré sur un ou plusieurs autres Pétris en les inversant à leur tour sur le velours.

3.5 Vérification de la capacité respiratoire des levures

Pour vérifier la capacité respiratoire de la souche de levure dépendante de RoBP1, on l'étale sur un milieu dont la seule source de carbone est le glycérol, qui est non-fermentable. Le milieu utilisé (YEPG) est composé de 3% de glycérol comme source de carbone, 10g/L de yeast extract, 20g/L de bacto-peptone, et 1,5% de bacto-agar. L'absence de croissance sur ce milieu après 5 jours d'incubation à 30°C indique un défaut au niveau de l'activité mitochondriale.

3.6 Rétrocroisement de la souche de levure dépendante de RoBP1

3.6.1 Croisement des levures

Une colonie fraîche de la souche dépendante de RoBP1 est mélangée sur gélose YEPD avec une colonie fraîche de la souche BY4705 MATa. Après une incubation à 30°C pendant toute la nuit, une gélose Yc-Ura est striée par épuisement à partir du mélange des deux souches et incubée pendant 48h à 30°C.

3.6.2 Identification des diploïdes

La ploïdie des colonies obtenues est confirmée par croisement avec les souches-test haploïdes Pta (MATa) et Ptα (MATα). Ces souches-test possèdent les marqueurs nutritionnels qui sont absents chez les 2 souches accouplées précédemment (ADE2, HIS3, LEU2, LYS2, MET15, TRP1 et URA3), mais sont auxotrophes pour l'isoleucine. Les colonies de levures diploïdes, incapables de s'accoupler avec les souches-test, ne peuvent pas acquérir les marqueurs et ne pourront donc pas croître sur un milieu minimal. De façon détaillée, une culture de 5 ml de chaque souche-test estensemencée à l'aide de colonies fraîches et incubée pendant toute la nuit à 30°C. Un volume de 150 µl de culture de chaque souche-test est étalé sur la surface de deux géloses YEPD. Les 2 Pétris sont incubés à 30°C pendant toute la nuit pour obtenir un tapis de levures. Les colonies obtenues après le premier croisement sont à leur tour croisées par « replica plating » sur un milieu YEPD avec les tapis de souches-test Pta

et Pta. Après une incubation pendant toute la nuit à 30°C, les Pétris sont transférés sur un milieu minimal et l'incubation est poursuivie pendant une autre nuit. Les colonies incapables de croître sur ce milieu sont diploïdes. Les colonies apparaissant sur ce milieu minimal sont des levures haploïdes ne s'étant pas accouplées initialement.

3.6.3 Sporulation des levures

Pour récupérer une levure haploïde dépendante de RoBP1, il s'agit de faire sporuler les levures diploïdes obtenues et de sélectionner les spores ayant conservé la dépendance. Il est possible d'induire la méiose et la formation de spores chez les levures diploïdes en les faisant croître dans un milieu pauvre en azote et en carbone. Un volume de 5 ml de milieu YEPD estensemencé avec une colonie fraîche de la souche de levure à faire sporuler. La culture est incubée avec agitation à 30°C jusqu'à saturation, correspondant à une concentration d'environ 2×10^8 cellules/ml. Un volume de 500 µl de cette culture (1×10^8 cellules) est transféré dans un autre tube et 5 ml d'eau stérile y sont ajoutés. Après centrifugation, le culot de levures est lavé à 2 autres reprises avec 5 ml d'eau stérile. Les levures sont resuspendues dans 5 ml de milieu de sporulation qui est constitué de 0,5% d'acétate de potassium et de 4 µg/ml des acides aminés suivants : histidine, leucine, lysine, méthionine et tryptophane. La culture est incubée avec agitation à 23°C pendant 5-7 jours et la présence de spores est vérifiée par observation directe sur une lame au microscope. Les 4 spores apparaissent alors groupées sous forme de tétrades contenues dans un asque.

3.6.4 Microdissection des spores

Les spores peuvent être récupérées individuellement par microdissection. Un volume de 1 ml de la culture de levure en sporulation est centrifugé dans un eppendorf stérile et le culot est lavé avec 1 ml d'eau stérile. Le culot de levures est resuspendu dans 50 μ l d'une solution de zymolyase permettant la digestion de la paroi cellulaire (5 μ l de zymolyase 0,5 mg/ml et 45 μ l de sorbitol 1 M). Les levures sont incubées dans cette solution pendant 10 min. à la température de la pièce et sont ensuite mises sur glace. Un volume de 20 μ l est ensuite déposé en une courte ligne à une extrémité d'une gélose YEPD. L'appareil MSM micromanipulator (Singer Instruments) est utilisé pour séparer les 4 spores issues d'une cellule diploïde et les disposer convenablement sur le Pétri (Figure 4). Le Pétri est ensuite incubé à 30°C jusqu'à ce que les spores forment des colonies.

3.6.5 Analyse des spores

Pour identifier le type sexuel des spores obtenues, on les croise sur milieu YEPD par « replica plating » avec les souches-test Pta et Pta α , comme décrit précédemment pour la détection de levures diploïdes après croisement. Les colonies issues de spores MATa s'accouplent seulement avec la souche Pta α et pourront de cette façon croître sur milieu minimal. La même analyse est faite avec les spores MATa. Les spores obtenues sont ensuite transférées sur des géloses Yc-Ura et 5'FOA par « replica plating » afin de sélectionner une souche toujours dépendante de RoBP1.

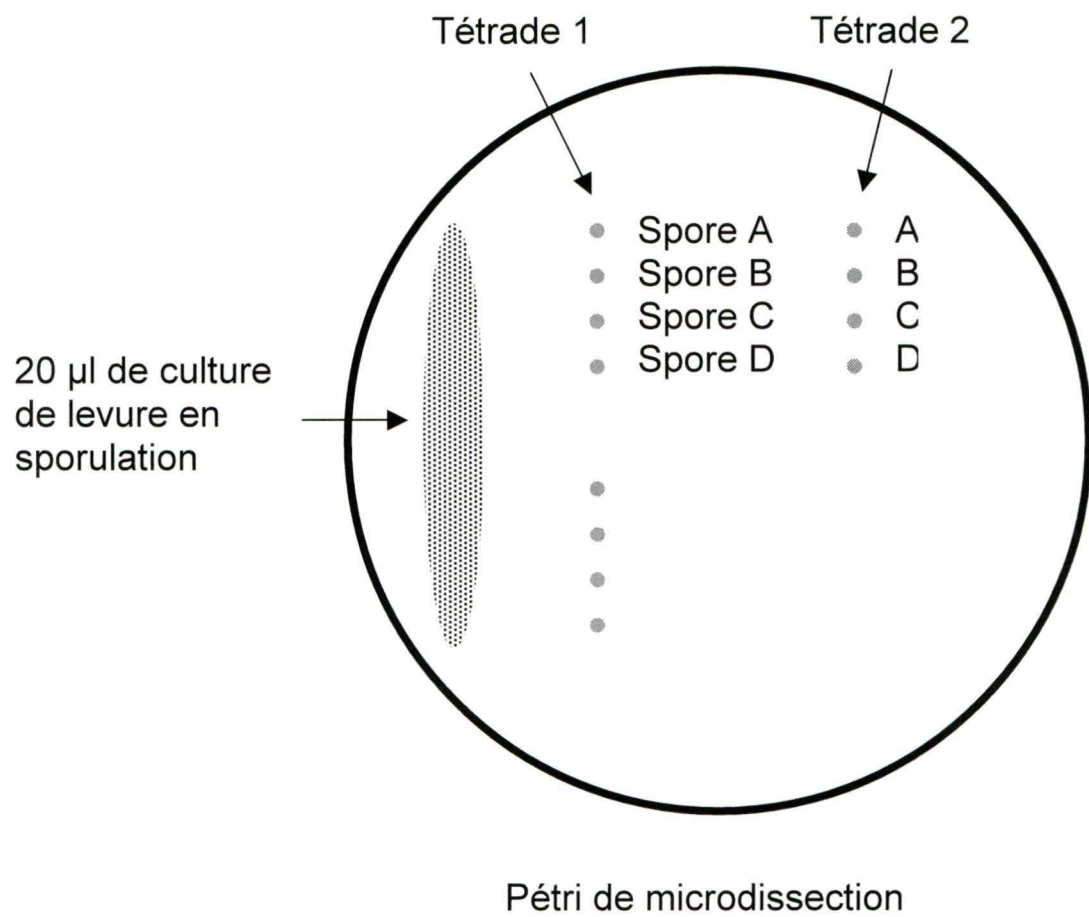


Figure 4 : Représentation de la méthode de microdissection des tétrades.

Les spores sont également transférées sur gélose YEPG pour vérifier leur capacité respiratoire.

3.7 Construction des plasmides pRS313/CAF4 et pRS313/UBC5

Deux des gènes susceptibles d'être les homologues de RoBP1 chez la levure, CAF4 et UBC5, ont été sous clonés dans le plasmide pRS313 à partir des plasmides pGAD424 Δ prom Δ GAL4/CAF4 (mémoire de Jim Boulanger) et pGAD424 Δ prom Δ GAL4/UBC5. Le plasmide pGAD424 Δ prom Δ GAL4/CAF4 est digéré séquentiellement avec les enzymes de restriction *EcoRI* et *SmaI* et le plasmide pGAD424 Δ prom Δ GAL4/UBC5 avec les enzymes *EcoRI* et *MluI*. Les fragments d'ADN correspondant à CAF4 et à UBC5 sont isolés par électrophorèse sur gel d'agarose et purifiés. Les extrémités du fragment UBC5, produites par les digestions enzymatiques, ont été enlevées pour créer des bouts francs en utilisant la Klenow (New England Biolabs) selon les recommandations du fabricant.

Le plasmide pRS313 est digéré séquentiellement avec les enzymes de restriction *EcoRI* et *SmaI*, d'une part, et seulement avec l'enzyme *SmaI*, d'autre part. Ces plasmides digérés sont isolés par électrophorèse sur gel d'agarose et purifiés. Les plasmides ainsi linéarisés sont ensuite traités à la phosphatase alcaline intestinale de veau. L'insert contenant CAF4 est ligué avec le plasmide pRS313 digéré *EcoRI/SmaI* et l'insert contenant UBC5 est ligué avec le plasmide pRS313 digéré *SmaI*. Les produits de ligation sont transformés en bactéries. Les plasmides issus des transformants obtenus sont digérés séquentiellement avec les enzymes de restriction

EcoRI et *SmaI*, pour confirmer la présence de l'insert CAF4, et avec les enzymes *EcoRI* et *BamHI*, pour confirmer la présence de l'insert UBC5. Les produits de digestion sont séparés par électrophorèse sur gel d'agarose pour vérifier leur taille.

3.8 Tests de complémentation

Afin de déterminer quels gènes de levure peuvent remplacer la fonction de RoBP1 dans la levure dépendante de RoBP1, des tests de complémentation sont effectués. Brièvement, le plasmide contenant le fragment d'ADN à tester est transformé dans la levure dépendante de RoBP1. Ce fragment d'ADN peut compléter pour la perte de RoBP1 s'il permet la croissance de la levure dépendante de RoBP1 sur un milieu 5'FOA qui force le rejet du plasmide contenant RoBP1.

3.8.1 Plasmides utilisés pour les tests de complémentation

Les plasmides utilisés dans les expériences de complémentation sont les plasmides pRS313, pRS313/UBC5, pRS313/CAF4, p366/chrIV, p366/chrXI, p366ΔUBC5, p366ΔchrIV, p366ΔchrXI et pGADGH/RoBP1. Les plasmides p366 et pGADGH possèdent le marqueur LEU2, tandis que le plasmide pRS313 possède le marqueur HIS3. Les plasmides p366/chrIV et p366/chrXI ont été récupérés de la bibliothèque génomique de *Saccharomyces cerevisiae* par complémentation (mémoire de Jim Boulanger). Les plasmides p366ΔUBC5, p366ΔchrIV et p366ΔchrXI ont été

construits en enlevant le gène UBC5 ou les fragments de chromosome IV ou XI des plasmides p366/chrIV et p366/chrXI. Le plasmide pGADGH contient l'ADNc de RoBP1 recueilli par RITA (BOUFFARD et al., 2000).

3.8.2 Transformation de plasmides dans les levures

La transformation de plasmides dans les levures est effectuée en utilisant une technique de perméabilisation de la paroi cellulaire avec l'acétate de lithium (GIETZ & WOODS, 2002). Une colonie fraîche sur Pétri de la levure dépendante de RoBP1 est piquée pour ensemer 2 ml de milieu de culture YEPD. La croissance des levures est poursuivie avec agitation à 30°C jusqu'à ce que la culture soit en phase stationnaire, à une densité optique de plus de 1,5 à une longueur d'onde de 600nm ($D.O_{600}$). Un volume de 0,1 ml de cette culture mère est utilisé pour ensemer 10 ml de milieu de culture YEPD. Lorsque cette nouvelle culture atteint une $D.O_{600}$ d'environ 0,680, soit une concentration d'environ 1×10^7 cellules/ml, le milieu de culture est retiré par centrifugation et le culot de levures est lavé dans 5 ml d'eau stérile. Le culot de levures est ensuite lavé dans un tube eppendorf avec 1 ml de Tris-EDTA (Tris-HCl 10 mM pH 8,0, EDTA 1 mM). Un troisième lavage est effectué avec 800 μ l de TE/LiAc (Tris-HCl 10 mM pH 8,0, EDTA 1 mM et LiAc 0,1 M) et le culot est resuspendu dans 50 μ l de TE/LiAc pour obtenir une concentration d'environ 2×10^9 cellules/ml. À ces levures maintenant compétentes sont ajoutés 1 μ g du plasmide à transformer, 50 μ g d'ADN de sperme de saumon préalablement bouilli et 300 μ l d'une solution de PEG4000 40% dans du TE/LiAc. Le mélange est incubé à

30°C avec agitation pendant 30 min. Un choc thermique est ensuite effectué à 42°C pendant 15 min. Les levures sont récupérées par centrifugation et le culot est resuspendu dans 1 ml de tampon TE. Les levures sont ensuite étalées sur des Pétris contenant les milieux restrictifs appropriés et la croissance est poursuivie à 30°C jusqu'à l'apparition des colonies de levures transformées (5-10 jours). Le maintien d'un plasmide dans une souche est assuré s'il fournit à la levure la capacité de produire un nutriment pour lequel elle est auxotrophe. De plus, la souche doit être cultivée sur un milieu dépourvu de ce nutriment pour assurer le maintien.

3.8.3 Croissance sur milieu 5'FOA

Les colonies transformées sont ensuite transférées par « replica plating » sur une gélose 5'FOA. Les Pétris sont incubés à 30°C jusqu'à l'apparition des colonies (5-10 jours). Les plasmides permettant la croissance sur ce milieu complètent pour la perte de RoBP1.

Le milieu 5'FOA est préparé en mélangeant 1,4g/L de la base azotée YNB sans sulfate d'ammonium ni acides aminés, 5g/L de sulfate d'ammonium, les quantités d'acides aminés pour obtenir les concentrations appropriées et 2% de bacto-agar (ROSE et al., 1990). Afin d'obtenir un pH final d'environ 6, il est nécessaire d'ajouter une goutte de NaOH 10N au mélange, car l'ajout de 5'FOA après l'autoclave acidifiera significativement le milieu. Après stérilisation à l'autoclave, le milieu est placé dans un bain à 65°C et, lorsqu'il est suffisamment refroidi, 2% de glucose filtré et 1g/L de 5'FOA (American Biorganics Inc.) y sont ajoutés. Afin de

bien dissoudre le 5'FOA, le milieu est agité pendant 45 min. à 65°C avant d'être coulé dans des boîtes de Pétri.

3.9 Création de levures dont les gènes CAF4 et/ou UBC5 ont été supprimés

Le gène UBC5 et le gène CAF4 sont supprimés individuellement dans la souche de levure BY4705. La souche contenant la délétion des 2 gènes à la fois a été créée à partir de la souche dont le gène UBC5 a été supprimé. La délétion des gènes est effectuée selon une méthode de transformation de fragments d'ADN générés par PCR (BAUDIN et al., 1993). Brièvement, un marqueur de sélection est amplifié par PCR en utilisant 2 oligonucléotides contenant chacun une portion homologue aux séquences situées de part et d'autre du gène à cibler. Lorsque transformé en levure, ce marqueur de sélection viendra remplacer le gène ciblé par recombinaison homologue.

3.9.1 Construction des fragments de délétion

Le marqueur de sélection LEU2 utilisé pour la délétion du gène UBC5 est amplifié par PCR à partir du plasmide pGAD424 avec les amorces UBC5-Leu-F (5'-TTG AAA AGT CAT TTT CTG CTC ACC ACC CTC AAC TAA ACT AAA AAT GTC TGC CCC TAA GAA-3') et UBC5-Leu-R (5'-GAG GAA GGT AAG TCT ACA CAA TTT ATC CGT TAG CCC AAA TTA ATC CCC TAC CCT ATG AAC-3'). La portion 3' de ces amorces sert à l'amplification de LEU2 et la portion 5' est constituée de séquences de 40 nucléotides homologues aux régions flanquant le gène

UBC5. Le marqueur de sélection TRP1 utilisé pour la délétion du gène CAF4 est amplifié par PCR à partir du plasmide pRS314 avec les amorces CAF4-Lys-R (5'-TGG AAA AAA AAA ATA GCT GTG CGG TAT TTC ACA CCG-3') et CAF4-Lys-F (5'-AAA AAA GTA AGA TTG TAC TGA GAG TGC AC-3'). La portion 3' de ces amorces sert à l'amplification de TRP1 et la portion 5' est constituée de séquences de 40 nucléotides homologues aux régions flanquant le gène CAF4. Le marqueur de sélection HIS3 utilisé pour la délétion du gène CAF4 dans la souche Δ UBC5 est amplifié par PCR à partir du plasmide pRS313 avec les amorces CAF4-Lys-R et CAF4-Lys-F. Les fragments de délétion sont isolés par électrophorèse sur gel d'agarose et purifiés.

3.9.2 Transformation de fragments de délétion et sélection des clones positifs

La transformation des fragments de délétion dans les levures est effectuée de la même façon que la transformation de plasmides décrite précédemment. Les fragments de délétion LEU2 et TRP1 sont transformés individuellement dans la levure BY4705 diploïde et le fragment de délétion HIS3 est transformé dans la levure diploïde dont le gène UBC5 a été supprimé précédemment. Après transformation, les levures sont étalées sur les milieux restrictifs appropriés (Yc-Leu, Yc-Trp ou Yc-Leu-His). La croissance est poursuivie à 30°C jusqu'à l'apparition des colonies.

3.9.3 Vérification des délétions

3.9.3.1 Préparation de l'ADN génomique des levures

L'extraction d'ADN génomique des levures est effectuée à l'aide d'une technique utilisant l'agitation de billes de verre pour briser physiquement les parois cellulaires (HUBERMAN et al., 1987; WELLINGER et al., 1993). Une colonie fraîche est piquée pour ensemercer 5 ml du milieu de culture YEPD. Après une croissance à 30°C pendant 16-20h, les levures sont lavées avec de l'eau stérile et resuspendues dans 500 µl de tampon de lyse (Tris 0,1 M, EDTA 50 mM et SDS 1%, pH 8,0). Le mélange est transféré dans un tube de verre de 13x100 mm. Des billes de verre préalablement lavées à l'acide nitrique sont ajoutées jusqu'à environ 2 mm du ménisque et le tout est conservé sur glace. Après avoir mélangé au vortex pendant 30 sec., 25 µl de NaCl 5 M sont ajoutés. L'agitation au vortex pendant 30 sec. est répétée 2 fois et le tube est ensuite laissé à la température de la pièce pendant 3 min. Le surnageant est transféré dans un tube eppendorf et 2 extractions successives avec 400 µl de phénol/chloroforme sont effectuées. Pour précipiter l'ADN, le surnageant est mélangé au vortex avec 1 ml d'éthanol 100% froid et laissé à -80°C pendant 20 min. L'ADN est précipité par centrifugation pendant 15 min. à 4°C. Le culot séché est resuspendu dans 50 µl de TE et 5 µl de RNase A concentrée à 5mg/ml et est incubé pendant 30 min. à 37°C. L'ADN est incubé pendant 1h à 37°C en présence de 150 µl de TNE (Tris 10 mM, NaCl 100 mM et EDTA 1 mM) et 5 µl de protéinase K concentrée à 10 mg/ml. Deux extractions successives avec 200 µl de

phénol/chloroforme et une avec 200 µl de chloroforme sont effectuées. L'ADN est mélangé avec 2 volumes d'éthanol 100%, incubé 20 min. à -80°C et précipité par centrifugation. Le culot d'ADN est resuspendu dans 50 µl de tampon TE.

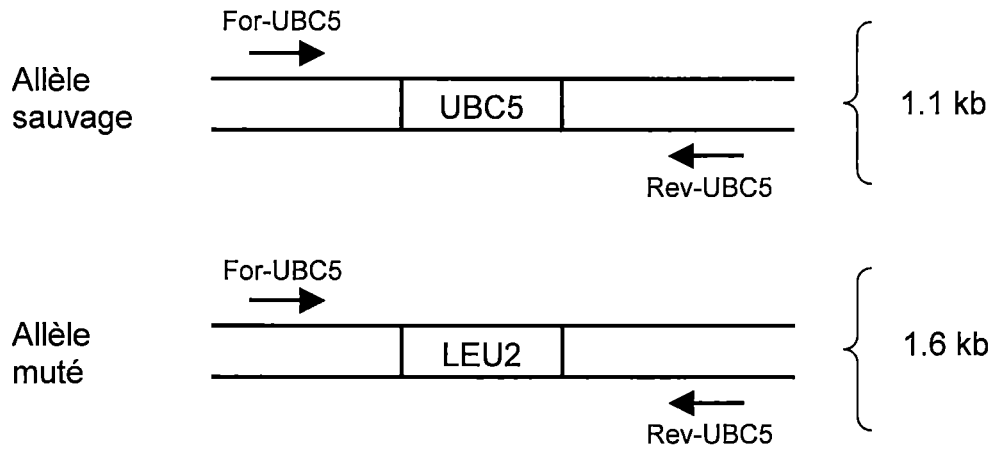
3.9.3.2 Vérification des délétions par PCR

L'insertion des marqueurs de sélection au niveau d'un des 2 allèles des gènes cibles est confirmée par PCR (Figures 5 et 6). La paire d'amorces utilisée pour la vérification de la délétion de CAF4 sont les amorces CAF4-3For (5'-CTG ACG AAA ATG GCA TGG-3') et CAF4-2Rev (5'-GGA CTG GAA TTC GAT ATT TCC TTT GTC CTA A-3'). La paire d'amorces utilisée pour la vérification de la délétion de UBC5 sont les amorces For-UBC5 (5'-TAA TGG TCA GAA TGG GCG CT-3') et Rev-UBC5 (5'-GGA GGT AGT TGC CAC TGC GA-3'). Les produits de PCR sont séparés par électrophorèse sur gel d'agarose et leur taille est analysée.

3.9.4 Sporulation, microdissection et analyse des spores

La sporulation des souches diploïdes dans lesquelles un des 2 allèles des gènes cibles est supprimé et la microdissection des tétrades sont effectuées tel que décrit précédemment. Pour identifier les levures haploïdes dans lesquelles les gènes UBC5 et/ou CAF4 sont remplacés par un marqueur de sélection, les spores sont transférées par « replica plating » sur les milieux restrictifs appropriés (Yc-Leu, Yc-Trp ou Yc-Leu-His). La croissance est poursuivie à 30°C jusqu'à l'apparition des colonies.

A



B

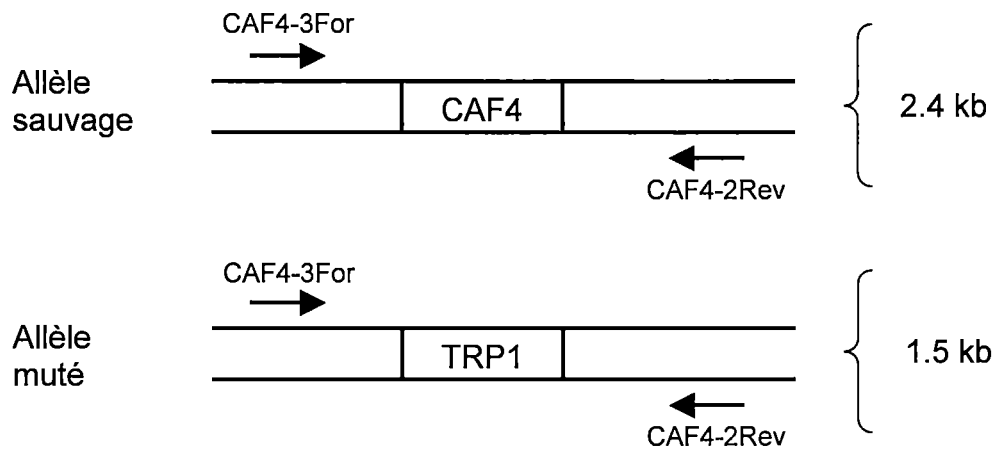


Figure 5 : Vérification des délétions simples UBC5 et CAF4 par PCR

Les flèches représentent les amorces utilisées pour le PCR. (A) Délétion UBC5. (B) Délétion CAF4.

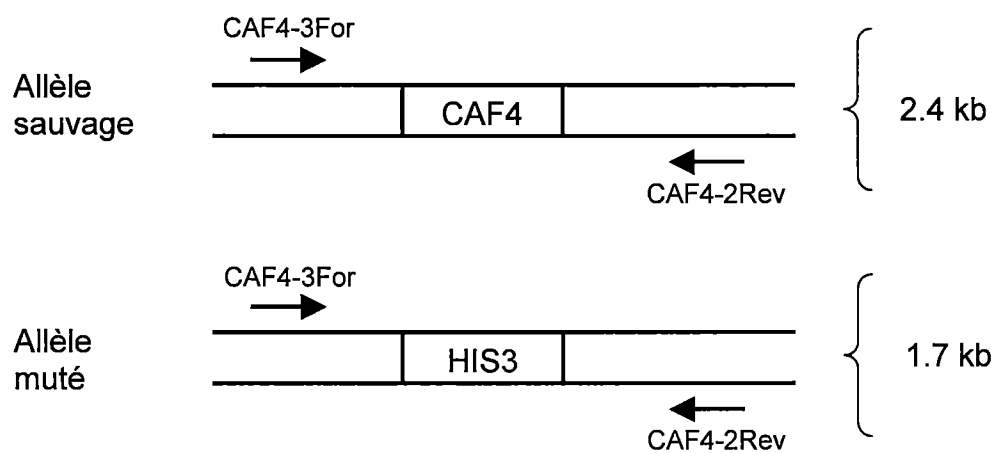


Figure 6 : Vérification de la délétion CAF4 dans la souche Δ UBC5 par PCR

Les flèches représentent les amorces utilisées pour le PCR.

Pseudogènes hY

3.10 Recherche de séquences homologues

Les séquences homologues aux gènes codant pour les 4 ARNs hY sont recherchées dans le génome humain (NCBI build 34.1) en utilisant l'outil MegaBLAST (ZHANG et al., 2000) disponible sur le site internet de NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>). La recherche est effectuée chez l'organisme *Homo sapiens* dans la base de données comprenant uniquement les séquences des chromosomes assemblés. Afin de maximiser le nombre de résultats, le paramètre « word size » est établi à 11 et le nombre de descriptions est établi à 1000. Les autres paramètres utilisés sont les paramètres par défaut. Le filtre de faible complexité est utilisé, mais les répétitions humaines ne sont pas masquées pour éviter de filtrer les pseudogènes hY.

Les résultats obtenus sont ensuite filtrés selon 2 critères de sélection de taille et d'homologie très stricts, pour limiter au maximum la probabilité d'obtenir une séquence simplement par chance (faux positif). De façon plus précise, seules les séquences étant à la fois (1) d'une taille supérieure ou égale à 50% de la taille de l'ARN hY correspondant (i.e $\geq 42-56$ nt) et (2) plus de 75% identiques à la séquence de l'ARN hY correspondant sont conservées pour être analysées. La même procédure est utilisée pour la recherche de séquences homologues à d'autres petits ARN non-codants. Sachant qu'il n'y a qu'une chance sur 300 d'avoir 2 séquences aléatoires de

20 nt identiques dans le génome humain, il se peut que les critères stringents utilisés aient menés à une sous-estimation du nombre de séquences dérivées des ARNs hY.

L'outil de recherche BLAT (KENT, 2002), disponible sur internet (<http://genome.cse.ucsc.edu/cgi-bin/hgBlat>), est utilisé pour la comparaison des pseudogènes Y entre les différentes espèces. La séquence des pseudogènes hY trouvés dans le génome humain, ainsi que les 500 nucléotides flanquants, est récupérée et utilisée pour la recherche de séquences homologues dans le génome du chimpanzé (*Pan troglodytes*; NCBI build 1 version 1).

3.11 Analyse de variation de séquence

Les pseudogènes hY retrouvés dans le génome humain sont alignés séquentiellement avec la séquence du gène de l'ARN hY correspondant en utilisant l'outil Matcher (<http://www.sanger.ac.uk>). Les variations entre les séquences des pseudogènes et des ARNs hY sont analysées et le pourcentage de mutation pour une position donnée est calculé. Les positions dont le pourcentage de variation dévie par plus de 2 fois l'écart-type du pourcentage moyen de variation sont analysées de façon plus approfondie. Pour cette statistique, seulement les pseudogènes contenant un nucléotide à cette position spécifique sont considérés. Par exemple, un pseudogène tronqué de 12 nucléotides à son extrémité 5' n'est pas utilisé pour les statistiques de la position 8.

3.12 Contenu en GC

La séquence des 5 kb flanquant chaque pseudogène hY trouvé est récupérée et le pourcentage des séquences en GC est calculé. La composition en GC des pseudogènes est également déterminée.

3.13 Localisation des pseudogènes

Les pseudogènes hY sont localisés dans le génome humain en utilisant ENSEMBL ([http://www.ENSEMBL.org/Homo sapiens](http://www.ENSEMBL.org/Homo_sapiens); version 18.34.1). Les positions de chaque pseudogène hY sur les chromosomes, attribuées par la recherche avec l'outil megaBLAST, sont utilisées pour rechercher des éléments caractéristiques dans la base de données de ENSEMBL. Les éléments caractéristiques récupérés sont classifiés en 3 groupes : exon, intron et intergénique. L'orientation des pseudogènes hY trouvés dans des gènes, comparativement à l'orientation du gène lui-même, est déterminée manuellement.

3.14 Analyse des signatures de rétrotransposition

En se basant sur les signatures caractéristiques laissées par une rétrotransposition par L1 (Figure 16), un programme (RTAnalyser) créé par Jean-François Lucier évalue la probabilité de rétrotransposition par la machinerie L1 pour un pseudogène donné. Cette probabilité est exprimée par le RetroScore. Après que

les extrémités 5' et 3' des pseudogènes hY soient correctement identifiées, le programme RTAnalyser utilise Matcher (<http://www.sanger.ac.uk>) pour identifier les TSDs de chaque côté des pseudogènes. Une queue poly(A) est également recherchée entre l'extrémité 3' du pseudogène hY et le TSD en 3'. Le site consensus de coupure par l'endonucléase L1 est recherché à l'extrémité 5' du pseudogène hY, avec 4 nucléotides chevauchant le TSD. Le poids des TSDs dans le RetroScore est le plus important (60% du maximum) car ils représentent le trait le plus caractéristique des éléments rétrotransposés par L1. La queue poly(A) compte pour 30% du RetroScore, car bien qu'elle soit souvent retrouvée dans la signature de rétrotransposition par L1, elle tend à raccourcir avec le temps. De plus, les séries d'adénines sont courantes dans le génome. Le site consensus de coupure par l'endonucléase L1 représente seulement 10% du RetroScore car il s'agit d'une courte séquence relativement peu conservée. Dans certains cas où les TSDs potentiels sont riches en séquences répétitives, des ajustements manuels sont requis.

La formule utilisée pour calculer le RetroScore est divisée en 3 parties, chacune mesurant séparément le pointage de la queue poly(A), des TSDs et du site de coupure consensus. Le RetroScore est calculé ainsi : $[A (5\sqrt{L^A}) - D^A] + [(100L^{\text{TSD}})/15 - (13M + 26G) - (10\sqrt{{}_5D^{\text{TSD}}}) - (2{}_3D^{\text{TSD}})] + (10S_{\text{freq}} - D^{\text{T}})$. A représente la proportion d'adénines dans la queue poly(A), L^A la taille de la queue poly(A) et D^A la distance (en nucléotides) entre l'extrémité 3' du pseudogène hY et la queue poly(A). L^{TSD} représente la taille des TSDs, M le nombre de mésappariements et G le nombre d'espacements (gap) entre les 2 TSDs, ${}_5D^{\text{TSD}}$ est la distance entre l'extrémité 5' du pseudogène hY et le TSD, et ${}_3D^{\text{TSD}}$ la distance entre l'extrémité 3' de la queue

poly(A) et le TSD. S_{freq} est la fréquence relative d'une séquence donnée comme site cible de l'endonucléase L1 (JURKA, 1997) et D^T la distance entre l'extrémité 5' du TSD et le site cible, considérant une superposition de 4 nucléotides. Le pointage minimum pour chacune des 3 mesures est zéro. Les séquences poly(A) situées à plus de 15 nucléotides du pseudogène hY ne sont pas considérées. Les multiplicateurs sont déterminés arbitrairement en fonction de l'importance relative de chaque variable et ajustés par des essais empiriques pour représenter le plus fidèlement ce qu'est un vrai événement de rétrotransposition. Par contre, le RetroScore calculé pour certains événements de rétrotransposition réels présentant des TSDs très courts ou très divergents est relativement faible. Par exemple, seulement 70% d'une sélection aléatoire d'éléments Alu obtient un RetroScore supérieur à 40.

Le pourcentage des pseudogènes hY tronqués dans leur portion 5' est calculé. Seulement les pseudogènes manquant au moins 10% de la séquence hY en 5' sont considérés comme tronqués.

3.15 Association avec des éléments Alu

Afin de vérifier s'il existe des pseudogènes hY chimériques, 500 nucléotides de chaque côté des pseudogènes hY sont séquentiellement alignés avec les séquences de l'élément Alu et de plusieurs autres petits ARNs non-codants (ARNt, snARN U1-6, snoARN, micro ARN, ARNr). Occasionnellement, des séquences Alu sont suffisamment près des pseudogènes hY pour déplacer un TSD à l'extérieur de la fenêtre utilisée pour la première analyse. Dans ces cas, l'analyse des signatures de

rétrotransposition est corrigée en tenant compte de cette séquence supplémentaire. L'outil RepeatMasker (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>) est également utilisé pour cette étude.

hY et épissage

3.16. Construction des plasmides

3.16.1 Modèle d'épissage alternatif hnRNP A1

Le plasmide 45, fourni gracieusement par le Dr Benoit Chabot (Département de microbiologie, Université de Sherbrooke), contient l'unité d'épissage alternatif dérivée du gène hnRNP A1 (Figure 7). Pour la construction des plasmides 45hY3 et 45hY3i (Figure 8), le plasmide 45 a été digéré par l'enzyme de restriction *EcoRV*, et pour la construction des plasmides 45hY3.1 et 45hY3.1i (Figure 8), le plasmide 45 a été digéré par l'enzyme de restriction *BamHI*. Les plasmides 45 ainsi linéarisés ont été isolés par électrophorèse sur gel d'agarose et purifiés. Les plasmides ont ensuite été traités à la phosphatase alcaline intestinale de veau afin d'éviter les faux positifs (ligation des deux extrémités libres d'un plasmide sans insert).

Les séquences d'ADN correspondant aux ARNs hY3 utilisées proviennent des plasmides pUC19hY3. Les séquences hY3 sont amplifiées par PCR avec des amorces contenant des sites de restriction pour les clonages. Les amorces SmaI-hY3-For (5'-AAA TCC CCC GGG CTG GTC CGA GTG CAG-3') et SmaI-hY3-Rev (5'-TCC

CCC GGG AAA AAG GCT AGT CAA GTG-3') ont été utilisées pour la construction des plasmides 45hY3 et 45hY3i. Les amorces BamHI-hY3-For (5'-ACG GGA TCC GGC TGG TCC GAG TGC-3') et BamHI-hY3-Rev (5'-ACG GGA TCC AAA GGC TAG TCA AGT GAA GC-3') ont été utilisées pour la construction des plasmides 45hY3.1 et 45hY3.1i. Les produits de PCR sont ensuite digérés avec les enzymes de restriction appropriées et les inserts sont isolés par électrophorèse sur gel d'agarose et purifiés. Les inserts sont finalement ligués avec les plasmides préparés et le tout est transformé en bactéries.

Pour confirmer la présence des inserts, les plasmides issus des transformants obtenus sont digérés avec l'enzyme de restriction *XhoI* et les produits de digestion sont séparés par électrophorèse sur gel d'agarose pour vérifier leur taille. L'orientation des inserts est vérifiée par une stratégie utilisant 2 réactions de PCR. Le PCR1 est effectué avec l'amorce E-Ad (5'-GAG TTT GTC CTC AAC CGC GA-3') et l'amorce en 5' utilisée pour la production de l'insert (ex : SmaI-hY3-For). Le PCR2 est effectué avec l'amorce E-Ad et l'amorce en 3' utilisée pour la production de l'insert (ex : SmaI-hY3-Rev). Les produits des PCR sont ensuite visualisés sur gel d'agarose. L'orientation des amorces est telle qu'un seul des 2 PCR doit donner un produit. L'insert est dans la même orientation que le mini-gène si le PCR1 donne un produit, et l'insert est dans l'orientation opposée si c'est le PCR2 qui donne un produit.

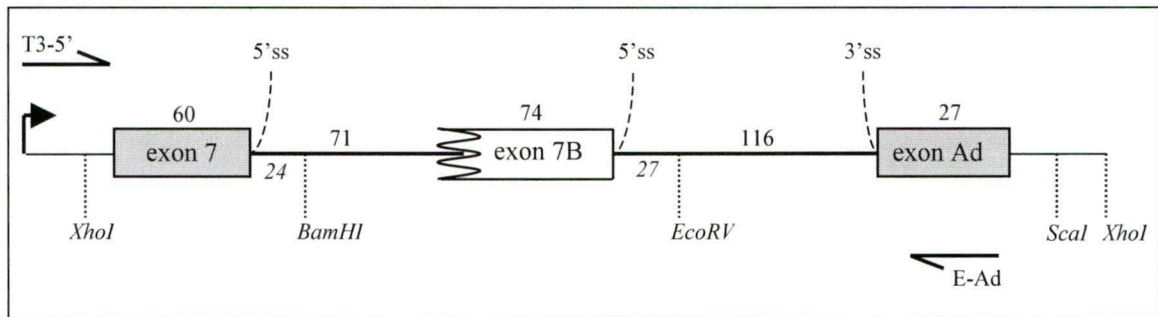


Figure 7 : Schéma de l'unité d'épissage alternatif dérivée du gène hnRNPA1.

Les 2 exons constitutifs sont représentés par des boîtes grises, l'exon alternatif par une boîte blanche et les 2 introns par des traits ondulés. La position des 2 sites d'épissage 5' (5'ss) et du site d'épissage 3' (3'ss) est indiquée au moyen de traits pointillés. Dans ce modèle, le site d'épissage 3' du premier intron est rendu inutilisable et est illustré ici par un trait ondulé vertical. La flèche coudée représente le site de début de transcription par l'ARN polymérase du phage T3. Les demi-flèches représentent la position des amorces utilisées pour les réactions de RT-PCR. La position des sites de restriction d'intérêt sur le mini-gène est indiquée à l'aide de traits pointillés. Les distances entre les sites de restrictions introniques utilisés pour les clonages (*BamHI* et *EcoRV*) et les sites d'épissage 5' sont indiquées en italique au-dessous du mini-gène. La taille des exons et introns est indiquée au-dessus de chaque élément.

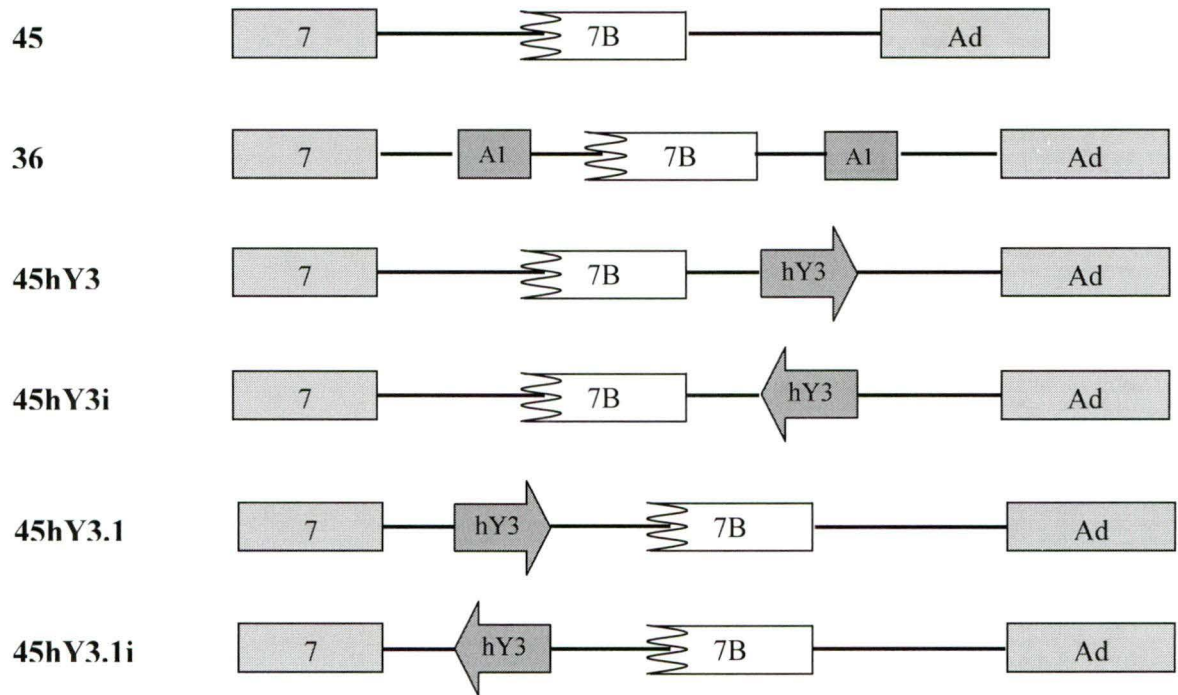


Figure 8 : Schémas et nomenclature des constructions utilisées dans les réactions d'épissage *in vitro* du mini-gène hnRNP A1.

Les boîtes A1 représentent des site de liaison pour la protéine hnRNP A1. Les flèches hY3 représentent la séquence d'ADN correspondant à l'ARN hY3. L'orientation des flèches représente l'orientation des séquences hY3 par rapport au mini-gène hnRNP A1. Les autres éléments des schémas sont décrits à la figure 7.

3.16.2 Modèle d'épissage alternatif Bcl-x

Le plasmide Bclx, fourni gracieusement par le Dr Benoit Chabot (Département de microbiologie, Université de Sherbrooke), contient l'unité d'épissage alternatif dérivée du gène Bcl-x (Figure 9). Pour la construction des plasmides Bclx3.1 et Bclx3.1i (Figure 10), le plasmide Bclx a été digéré simultanément avec les enzymes de restriction *BsmBI* et *BstEII*. Les produits de digestion ont ensuite été séparés par électrophorèse sur gel d'agarose pour isoler le plasmide du fragment *BsmBI* / *BstEII*. La bande d'ADN correspondant au plasmide a été excisée et purifiée du gel. Les extrémités laissées par les digestions enzymatiques ont été enlevées pour créer des bouts francs en utilisant la Klenow (New England Biolabs) selon les recommandations du fabricant. Le plasmide a été ensuite traité à la phosphatase alcaline intestinale de veau pour éviter la ligation du plasmide sur lui-même. Pour la construction des plasmides Bclx3.2 et Bclx3.2i (Figure 10), la procédure est la même que celle utilisée pour les plasmides Bclx3.1 et Bclx3.1i, excepté que les enzymes de restriction utilisées sont *BsaBI* et *BseRI*.

La séquence d'ADN correspondant à l'ARN hY3 est amplifiée par PCR à partir du plasmide pUC19hY3 avec les amorces EcoRI-hY3-For (5'-ACG GAA TTC GGC TGG TCC GAG TGC-3') et BamHI-hY3-Rev, qui contiennent des sites de restriction pour les clonages. Pour permettre la ligation avec les extrémités déphosphorylées du plasmide, la *T4 Polynucleotide Kinase* (New England Biolabs) est utilisée pour ajouter un phosphate aux extrémités 5' des inserts produits par PCR. Les inserts sont ensuite ligués avec les plasmides et le tout est transformé en bactéries.

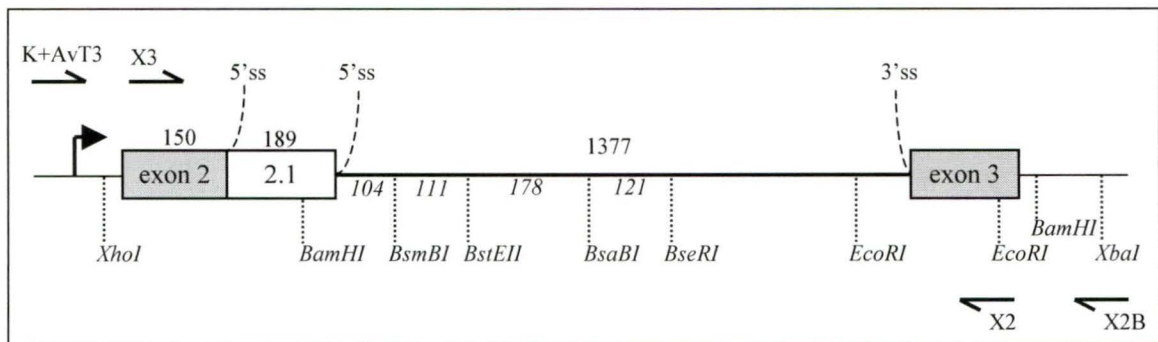


Figure 9 : Schéma de l'unité d'épissage alternatif dérivée du gène Bcl-x.

Les séquences exoniques constitutives sont représentées par des boîtes grises, la séquence alternative par une boîte blanche et l'intron par un trait foncé. La position des 2 sites d'épissage 5' (5'ss) et du site d'épissage 3' (3'ss) est indiquée au moyen de traits pointillés. La flèche coudée représente le site de début de transcription utilisé dans les expériences *in vitro*. Les demi-flèches représentent la position des amorces utilisées pour les réactions de transcription inverse et de PCR. La position des sites de restriction d'intérêt sur le mini-gène est indiquée à l'aide de traits pointillés. Les distances entre les sites de restriction introniques utilisés pour les clonages sont indiquées en italique au-dessous du mini-gène. La taille des exons et de l'intron est indiquée au-dessus de chaque élément. Le schéma est aussi valable pour les expériences *in vivo*, sauf que le fragment *XhoI-XbaI* est cloné devant un promoteur CMV dans le vecteur pBluescript II KS(+).

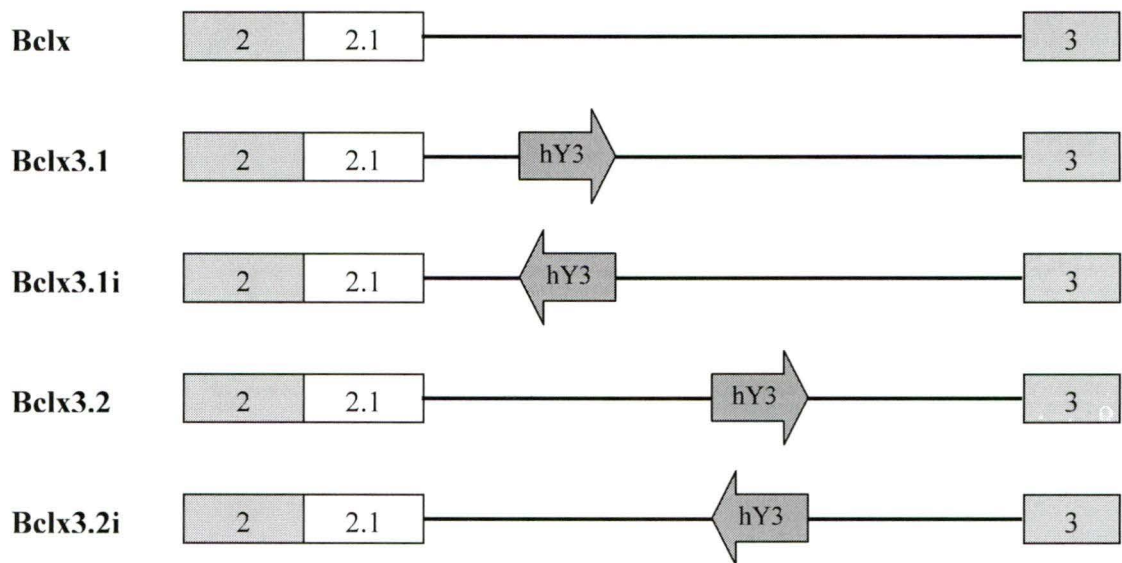


Figure 10 : Schémas et nomenclature des constructions utilisées dans les réactions d'épissage *in vitro* et *in vivo* du mini-gène Bcl-x.

Les flèches hY3 représentent les séquences d'ADN ajoutées correspondant à l'ARN hY3. L'orientation des flèches représente l'orientation des séquences hY3 par rapport au mini-gène Bcl-x. Les autres éléments des schémas sont décrits à la figure 9.

Pour confirmer la présence et l'orientation des inserts, les plasmides issus des transformants obtenus sont digérés avec l'enzyme de restriction *BamHI* et les produits de digestion sont séparés par électrophorèse sur gel d'agarose pour vérifier leur taille. L'obtention de seulement 2 produits de digestion indique l'absence d'insert. La présence de l'insert est confirmée lorsqu'on obtient 3 produits de digestion. L'orientation des inserts par rapport au mini-gène est déterminée en analysant les tailles des fragments obtenus.

Pour la construction des plasmides utilisés pour les expériences d'épissage *in vivo*, les plasmides Bclx, Bclx3.1, Bclx3.1i, Bclx3.2 et Bclx3.2i sont digérés avec l'enzyme de restriction *XhoI*. Les plasmides digérés sont traités à la Klenow pour rendre les extrémités franches, pour être ensuite digérés avec l'enzyme de restriction *XbaI*. Le fragment *XhoI* / *XbaI* est ensuite purifié sur gel d'agarose. Le plasmide pcDNA3.1+ est digéré avec les enzymes de restriction *XbaI* et *EcoRV* et purifié sur gel d'agarose. Le plasmide pcDNA3.1+ digéré est ligué individuellement avec les fragments récupérés des plasmides Bclx, Bclx3.1, Bclx3.1i, Bclx3.2 et Bclx3.2i, pour être ensuite transformé en bactéries.

3.17 Transcription des ARNs pré-messagers pour l'épissage *in vitro*

3.17.1 Préparation des matrices

3.17.1.1 Modèle hnRNP A1

Les ARNs pré-messagers utilisés lors des expériences d'épissage *in vitro* sont transcrits *in vitro* à partir d'une matrice d'ADN. Chaque unité d'épissage alternatif construite est sous le contrôle du promoteur de l'ARN polymérase du phage T3. Comme l'unité ne possède pas de signal de terminaison de la transcription, il est nécessaire de modifier la matrice de façon à ce que la transcription s'arrête à l'extrémité 3' de l'unité d'épissage. Pour la préparation des matrices utilisées dans les expériences d'épissage du modèle hnRNP A1, 10-15 µg du plasmide d'intérêt sont digérés avec l'enzyme de restriction *ScaI*. L'état de la digestion est observé sur gel d'agarose et si elle est incomplète, de l'enzyme *ScaI* est ajouté jusqu'à ce que la digestion soit complète. Pour éliminer les RNases et les enzymes de restriction, le plasmide linéarisé est incubé avec 1% de *Proteinase K* (Amersham Biosciences), 0,2% de SDS et 1mM de CaCl₂ pendant 30 min. à 37°C. L'inactivation de la protéinase K est faite par une incubation à 68°C pendant 10 min. Le plasmide est ensuite mélangé avec 400µl de phénol/chloroforme/alcool isoamyl (25:24:1) et 0,1M de NaCl à l'aide d'un vortex et l'ADN présent dans la phase aqueuse est ensuite précipité avec 1ml d'éthanol 100%. Le culot d'ADN est resuspendu dans 5µl d'eau, pour obtenir une concentration d'environ 1µg/µl.

3.17.1.2 Modèle Bcl-x

La préparation des matrices pour la transcription des ARNs pré-messagers utilisés dans les expériences d'épissage *in vitro* du modèle Bcl-x a été faite par PCR. Les réactions de PCR utilisent les amorces nommées K+AvT3 (5'-GAC CAT GAT TAC GCC AAG CG-3') et X2B (5'-CGC TCT AGA ACT AGT GGA TC-3'), l'enzyme *PfuTurbo DNA polymerase* (Stratagene) et les plasmides appropriés. La procédure utilisée est la suivante : 95°C pendant 2 min.; 30 cycles à 95°C pour 1 min., 55°C pour 55 sec. et 72°C pour 2 min.; et une extension finale à 72°C pendant 10 min. Les produits de PCR sont ensuite purifiés sur gel d'agarose.

3.17.2 Réaction de transcription des ARNs pré-messagers

Les ARNs pré-messagers sont transcrits à partir des matrices d'ADN préparées préalablement. Pour la transcription des ARNs pré-messagers du modèle hnRNP A1, 5µl de matrice à 1µg/µl sont mélangés avec 20µl de SP6 Buffer 5X, 10µl d'UTP à 5mM, 8µl de nucléotides analogue du capuchon m7G(5')ppp(5')G, 2µl de l'inhibiteur de ribonucléase *RNAguardTM*, 10µl de rNTP à des proportions de concentrations de ATP : CTP : GTP à 5 : 5 : 1 mM, 0,5 ul de [α -³²P]UTP (Perkin Elmer ; 800 Ci/ml), 42,5µl d'eau et finalement 2µl d'ARN polymérase du phage T3 (USB), ce qui donne un volume total de 100µl. Le fournisseur de ces réactifs est Amersham Biosciences lorsque non mentionné. Le tout est incubé à 37°C pendant une heure. La réaction de transcription est arrêtée par l'ajout de 50µl de tampon de migration contenant 95% de

formamide et du bleu de bromophénol. La réaction de synthèse des transcrits pour le modèle Bcl-x est effectuée de la même manière, sauf que la réaction utilise 47,5µl de matrice à environ 100 ng/µl pour compléter le volume à 100µl.

3.17.3 Purification et dosage des ARNs transcrits

Les ARNs transcrits sont purifiés par électrophorèse sur gel de polyacrylamide. Les échantillons contenant les transcrits sont incubés pendant 1 min. 30 sec. dans de l'eau bouillante, puis laissés sur glace pendant 1 min. 30 sec. Les échantillons sont ensuite séparés par électrophorèse dans un gel formé de 5% de polyacrylamide : bisacrylamide (38 : 2), de TBE (90mM Tris-borate, 1mM EDTA) et de 7M d'urée pendant 30-45 min. à 350V. Le gel est démoulé puis exposé 45 min. sur film d'autoradiographie *BioMax XAR* (Kodak) pour visualiser les bandes. Les bandes radioactives correspondant aux ARNs transcrits sont découpées et déposées dans des tubes. La méthode du "crush and soak" (CHABOT, 1994) est utilisée pour éluer l'ARN du gel. Les bandes de gel sont écrasées jusqu'à ce qu'elles aient une apparence visqueuse, puis 500 µl de solution d'élution composée de 0,3M de NaOAc à pH 5,2 et de 0,2% de SDS est ajouté au gel. Le tout est incubé sous rotation pendant 1 heure à la température de la pièce. Le surnageant est récupéré après 5 min. de centrifugation et soumis à 2 extractions au phénol/chloroforme/alcool isoamyl successives. L'ARN présent dans la phase aqueuse est précipité avec 1ml d'éthanol 100%. La quantité de radioactivité présente dans le culot est mesurée à l'aide de l'appareil compteur de radioisotopes de table *Quick-Count* (Bioscan inc.). La

concentration d'ARN transcrit est estimée en tenant compte de la quantité de radioactivité et du nombre d'uridine présentes dans l'ARN transcrit. Le culot est resuspendu dans l'eau pour obtenir une concentration approximative de 2 fmol/ μ l.

3.18 Épissage des ARNs pré-messagers

3.18.1 Réaction d'épissage *in vitro*

Une réaction d'épissage *in vitro* typique est composée de 0,5 μ l des réactifs suivants : créatine kinase à 2 U/ μ l, rATP à 12,5mM, $MgCl_2$ à 80 mM et créatine phosphate, de 0,25 μ l de DTT à 100 mM, de 0,25 μ l de RNA Guard, de 2,5 μ l d'alcool polyvinyle (PVA) à 13%, de 2,5 μ l de tampon D (60 mM HEPES pH 7,9, 100 mM KCl, 0,2 mM EDTA, 0,5 mM PMSF, 0,5 mM DTT, et 20% glycérol) et de 5 μ l d'extrait nucléaire de cellules HeLa (DIGNAM, 1983), gracieusement fourni par le laboratoire du Dr Chabot. Le mélange est incubé à 30°C pendant 2 heures en présence de 2 fmol de l'ARN transcrit approprié. La réaction d'épissage est arrêtée par l'addition de 450 μ l de tampon d'élution (0,3M de NaOAc à pH 5.2 et 0,2% de SDS) et 500 μ l de phénol/chloroforme/alcool isoamyl. Pour faciliter la précipitation de l'ARN, 3 μ l d'ARNt de levure est ajouté. Le tout est mélangé à l'aide d'un vortex et l'ARN présent dans la phase aqueuse est ensuite précipité avec 1ml d'éthanol 100%. Le culot est resuspendu dans l'eau à une concentration de 5 amol/ μ l.

3.18.2 Réaction d'épissage *in vitro* avec ARN hY compétiteur

Afin de déterminer si la modulation de l'épissage observée dépend de l'action de facteurs agissant en *trans*, un ARN compétiteur est ajouté à la réaction d'épissage *in vitro*. Les ARNs compétiteurs hY3 et hY3 antisens (hY3i) sont transcrits à partir des plasmides pUC19hY3 et pTZRhY3, respectivement. Pour s'assurer d'une terminaison adéquate de la transcription, les plasmides pUC19hY3 et pTZRhY3 sont préalablement digérés avec les enzymes de restriction *DraI* et *EcoRI*, respectivement. Les plasmides linéarisés sont ensuite purifiés sur gel d'agarose. La méthode de transcription et de purification des ARNs compétiteurs est la même que celle utilisée pour les ARNs pré-messagers, excepté que le nucléotide radiomarké est le CTP, que la polymérase utilisée est l'ARN polymérase du phage T7 et qu'il n'y a pas d'analogue du capuchon. Les ARNs compétiteurs sont finalement resuspendus dans l'eau à différentes concentrations.

Avant l'addition de l'ARN pré-messager au mélange pour la réaction d'épissage, 1 µl d'ARN compétiteur, à la concentration appropriée, est ajouté au mélange et préincubé pendant 10 min. à 30°C. Le mélange pour la réaction d'épissage a la même composition que décrite précédemment, excepté qu'il n'y a que 1,5 µl de tampon D. Ensuite, 1 µl d'ARN pré-messager est ajouté à la réaction et le tout est incubé à 30°C pendant 2 heures. Le reste de la procédure est la même que celle décrite précédemment pour l'épissage *in vitro*.

3.18.3 Réaction d'épissage *in vivo* (Bcl-x)

3.18.3.1 Transfection de cellules HeLa S3 et ECR 293

Les cellules sont cultivées dans le milieu Dulbecco's modified Eagle's contenant 10% de sérum fœtal bovin et 1% de glutamine. Pour les transfections, 2×10^5 cellules sont étalées dans des puits de 35 mm². Les transfections ont été réalisées en utilisant 1,5 µg de plasmide et 3 µl de Lipofectamine (Invitrogen), selon les recommandations du fabricant.

3.18.3.2 Extraction d'ARN

L'extraction de l'ARN total des cellules est effectuée 24 heures après transfection en utilisant le *TRIzol* (Invitrogen), selon les recommandations du fabricant. L'ARN est ensuite traité à la DNase I pendant 10 min. à 37°C, suivi d'un traitement à 70°C pendant 5 min. Un dixième de l'ARN extrait total est utilisé pour être rétrotranscrit avec la trousse *Omniscript RT* (Qiagen).

3.19 Détection et quantification des produits d'épissage

3.19.1 Modèle hnRNP A1

Afin d'être en mesure de détecter et de quantifier les différentes espèces

d'ARN résultant de l'épissage alternatif des ARN pré-messagers, les produits d'épissage sont amplifiés par une réaction de RT-PCR utilisant un nucléotide radiomarké et ensuite visualisés sur un gel de polyacrylamide. Pour les expériences utilisant le modèle hnRNP A1, l'amplification par RT-PCR est effectuée en une seule étape à l'aide de la trousse *Ready-To-GoTM RT-PCR Beads* (Amersham Biosciences). Les réactions de RT-PCR sont réalisées en utilisant 10 amoles d'ARN épissé comme matrice, 8 ng des amorces E-Ad et T3-5' (5'-AAT TAA CCC TCA CTA AAG GG-3'), 0,1 ul de [α -³²P]dCTP (Perkin Elmer ; 800 Ci/ml), et de l'eau pour obtenir un volume total de 50 μ l. La procédure utilisée est la suivante : transcription inverse à 42°C pendant 45 min.; 94°C pendant 2 min.; 30 cycles à 94°C pour 30 sec., 55°C pour 30 sec. et 72°C pour 20 sec.; et une extension finale à 72°C pendant 15 min.

Les produits de la réaction de RT-PCR sont séparés par électrophorèse dans un gel formé de 5% de polyacrylamide : bisacrylamide (38 : 2) et de TBE (90mM Tris-borate, 1mM EDTA) pendant 1 heure à 250V. Le gel est démoulé et les bandes sont visualisées directement en utilisant l'appareil Instant Imager (Canberra Packard), ou, après exposition sur écran photostimulable dans une cassette, à l'aide de l'appareil Phosphor Imager STORM 860 (Amersham Biosciences). Pour évaluer la modulation du choix entre les 2 sites d'épissage 5', l'intensité des bandes correspondant aux 2 produits d'épissage alternatif (proximal et distal) est mesurée et comparée à l'aide du logiciel approprié fourni avec les appareils.

3.19.2 Modèle Bcl-x

Pour les expériences utilisant le modèle Bcl-x, l'amplification par RT-PCR est effectuée en 2 étapes. Tout d'abord, les produits d'épissage sont rétrotranscrits en ADN à l'aide de la trousse *Omniscript RT* (Qiagen). La réaction est effectuée selon les recommandations du fabricant en utilisant l'amorce X2B et 10 amoles d'ARN épissé. Par la suite, 1 µl de la réaction de transcription inverse est utilisé comme matrice dans une réaction de PCR avec les amorces X3 (5'-ATG GCA GCA GTA AAG CAA GCG-3') et X2 (5'- TCA TTT CCG ACT GAA GAG TGA-3'), et contenant 0,1 µl de [α -³²P]dCTP. La procédure utilisée est la suivante : 95°C pendant 3 min.; 35 cycles à 93°C pour 15 sec., 60°C pour 30 sec. et 72°C pour 30 sec.; et une extension finale à 72°C pendant 3 min. Les produits de PCR sont séparés par électrophorèse dans un gel formé de 4% de polyacrylamide : bisacrylamide (29 : 1) et de TBE (90mM Tris-borate, 1mM EDTA) pendant 1 heure à 250V. La quantification des produits d'épissage est effectuée de la même façon qu'avec le modèle hnRNP A1. Pour l'épissage *in vivo*, l'ARN cellulaire extrait sert de matrice à la réaction de transcription inverse.

4. RÉSULTATS

RoBP1 et levure

4.1 Levure dépendante de RoBP1: confirmation de la dépendance.

La souche de levure dépendante de l'expression de la protéine RoBP1 sert de base à la recherche d'homologues fonctionnels de cette protéine humaine chez *Saccharomyces cerevisiae*. La première étape est donc la confirmation de la dépendance à RoBP1. La souche dépendante de RoBP1 utilisée par Jim Boulanger provient directement de la mutagenèse et est dite de 1^{ère} génération. La souche dépendante de RoBP1 dite de 2^e génération est issue d'un rétrocroisement de la souche de 1^{ère} génération avec la souche haploïde BY4705 MATa. La souche dépendante de RoBP1 utilisée dans les travaux présentés ici est la souche de 2^e génération, excepté lors de mention contraire.

Même après plusieurs semaines de croissance sur un milieu riche non-sélectif (YEPD), cette souche de levure produit des colonies uniformément blanches. Ce phénotype indique la conservation du marqueur de sélection ADE2, situé sur le plasmide pCu4X6/RoBP1, malgré l'absence de sélection. Sans pression de sélection, ce type de plasmide est perdu à un taux d'environ 3% par génération (DANI & ZAKIAN, 1983). De plus, la levure dépendante de RoBP1 réussit à croître lorsque transférée sur un milieu déficient en uracile, ce qui confirme la conservation du marqueur de sélection URA3 situé sur le plasmide pCu4X6/RoBP1.

Pour confirmer la dépendance à RoBP1, les levures sont transférées sur gélose 5'FOA. Seules les levures dépourvues du marqueur URA3 peuvent croître sur ce milieu. La souche de levure dépendante de RoBP1 est incapable de croître sur le milieu 5'FOA après 4 jours à 30°C. Cependant, on voit quelquefois apparaître des micro-colonies. Par comparaison, la souche contrôle BY4704 (URA3+) ne produit aucune colonie après 4 jours de croissance sur le milieu 5'FOA et la levure sauvage BY4705 (ura3-), montre une bonne croissance sur ce milieu après 4 jours. Malgré l'interprétation parfois ambiguë de la croissance, il semble donc que la levure dépendante de RoBP1 ne peut se départir du plasmide contenant RoBP1.

Il a été noté par contre que la souche dépendante de RoBP1 produit des colonies de petite taille sur gélose YEPD après 4 jours de croissance à 30°C, comparativement à la souche de levure BY4705 haploïde sauvage qui donne des colonies de taille significativement plus grande après seulement 2 jours (Figure 11).

4.2 Capacité respiratoire de la souche dépendante de RoBP1 : la souche dépendante de RoBP1 a un phénotype de type petite.

Certaines levures sont incapables d'utiliser les sucres en condition aérobie dû à un défaut au niveau de l'activité mitochondriale. Ces levures doivent plutôt fermenter les sucres. Comme la fermentation produit un rendement énergétique inférieur, ces levures, dites « petites », donnent sur milieu solide des petites colonies (FAYE et al., 1973). Pour vérifier si la croissance ralentie de la souche dépendante de RoBP1 est due à un problème mitochondrial, les levures sont transférées sur gélose YEPG. La

seule source de carbone de ce milieu est le glycérol qui est non-fermentable. Après 4 jours de croissance sur gélose YEPG, les levures sauvages BY4705 et BY4704 produisent de petites colonies, par contre, la levure dépendante de RoBP1 est incapable de croître (Figure 12). Donc cette souche possède des fonctions mitochondriales altérées.

4.3 Analyse du phénotype de toutes les souches de levures dépendantes de RoBP1

Le phénotype de toutes les souches de levure dépendante de RoBP1 disponibles a été observé sur les différents milieux (Tableau 3).

La croissance de la souche dépendante de RoBP1 créée par Jim Boulanger (1^{ère} génération) est similaire à celle de la souche dépendante de RoBP1 de 2^e génération : croissance sur Yc-Ura, mais aucune croissance sur YEPG et 5'FOA. La seule différence consiste en une croissance légèrement supérieure sur gélose YEPD. Donc, la souche dépendante de RoBP1 de 1^{ère} génération est également petite. Un premier rétrocroisement de cette souche dépendante de RoBP1 de 1^{ère} génération a été effectué précédemment avec la souche BY4705 sauvage. En plus de la souche utilisée jusqu'à maintenant dans ce projet, 5 autres spores issues de ce rétrocroisement (2^e génération) auraient aussi conservé la dépendance à RoBP1, selon Jim Boulanger. Les spores 1.9b, 1.1b, 3.3d et 2.8d montrent exactement le même phénotype que la souche dépendante de RoBP1 utilisée jusqu'à maintenant. Ces spores semblent dépendantes de RoBP1, mais sont petites. Par contre, la spore 5.8b montre un phénotype différent. Sa croissance sur milieu YEPD, comparable à la croissance de la

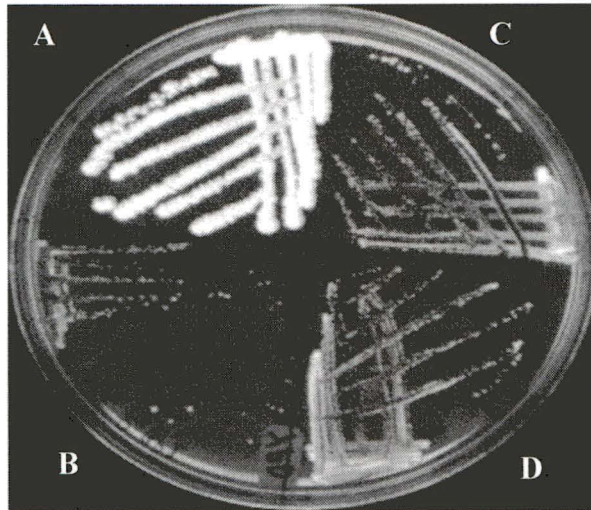


Figure 11 : La levure dépendante de RoBP1 donne de petites colonies sur milieu YEPD.

Croissance après 4 jours à 30°C. (A) Souche sauvage BY4705. (B, C et D) Souche dépendante de RoBP1.

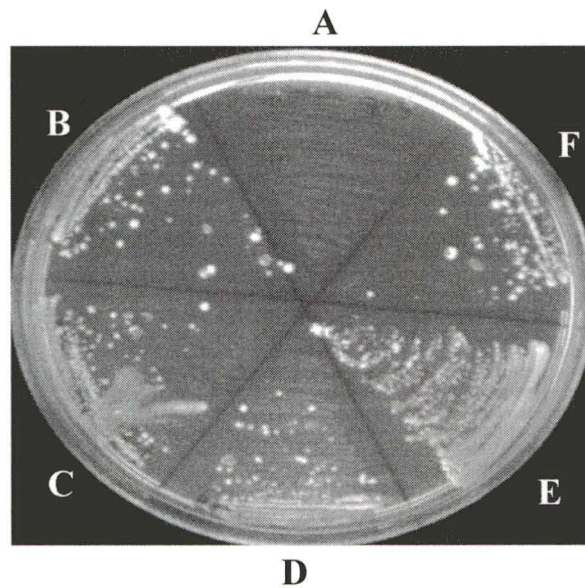


Figure 12 : La levure dépendante de RoBP1 ne pousse pas sur milieu YEPG.

Croissance après 4 jours à 30°C. (A) Souche dépendante de RoBP1. (B) Souche BY4705 diploïde. (C et D) Souche BY4705 haploïde. (E et F) Souche BY4704 haploïde.

souche BY4705 sauvage, et sa faible croissance sur milieu YEPG suggèrent des fonctions mitochondriales normales. La croissance de la spore 5.8b sur milieu Yc-Ura indique la présence du plasmide avec le marqueur URA3 (pCu4X6/RoBP1), mais sa légère croissance sur milieu 5'FOA montre que les levures peuvent se débarrasser du plasmide pour permettre une croissance. La spore 5.8b ne semble donc pas être petite, mais semble être peu dépendante de RoBP1. Étant donné ses phénotypes intermédiaires sur milieu YEPG et 5'FOA, la spore 5.8b n'a pas été étudiée plus en détail.

Donc, aucune souche ne montre une claire dépendance à RoBP1 tout en ayant des fonctions mitochondriales normales.

4.4 Rétrocroisement de la levure dépendante de RoBP1 avec une levure sauvage : la souche dépendante de RoBP1 reste petite.

Afin de récupérer une levure dépendante de RoBP1 dont les fonctions mitochondriales sont normales, un deuxième rétrocroisement de la levure dépendante de RoBP1 avec la levure BY4705 sauvage a été effectué. Après sporulation des levures diploïdes et microdissection des tétrades, la croissance des spores a été vérifiée sur milieu YEPD, Yc-Ura, 5'FOA et YEPG. Aucune des 60 tétrades disséquées ne produit de spores ayant conservé la dépendance à RoBP1 (croissance sur Yc-Ura et absence de croissance sur 5'FOA) et ayant récupéré des fonctions mitochondriales normales (croissance normale sur YEPD et croissance sur YEPG). Les spores dépendantes de RoBP1 sont toujours petites.

Souches	Milieu			
	YEPD	Yc-Ura	YEPG	5'FOA
Dép. RoBP1 1 ^{ère} génération	++	++	-	-
Dép. RoBP1 2 ^e génération	+	++	-	-
1.9b	+	++	-	-
1.1b	+	++	-	-
5.8b	+++	++	+	+
3.3d	+	++	-	-
2.8d	+	++	-	-
BY4705 sauvage (ura3-)	+++	-	++	++
BY4704 sauvage (URA3+)	+++	++	++	-

Tableau 3 : Phénotype des différentes souches de levure dépendantes de RoBP1 et des souches contrôles.

Taille des colonies après 4 jours à 30°C. (+++) grande taille, (++) taille moyenne, (+) petite taille, (-) aucune colonie. Les souches 1.9b, 1.1b, 5.8b, 3.3d et 2.8d sont d'autres spores dépendantes de RoBP1 de 2^e génération issues d'un rétrocroisement entre la souche de 1^{ère} génération et la souche BY4705 sauvage.

4.5 Tests de complémentation : aucune complémentation claire

Les plasmides contenant les gènes ou fragments d'ADN susceptibles de remplacer la fonction de RoBP1 sont transformés dans la levure dépendante de RoBP1 (pRS313, pRS313/UBC5, pRS313/CAF4, p366/chrIV, p366/chrXI, p366ΔUBC5, p366ΔchrIV, p366ΔchrXI et pGADGH/RoBP1). Ces plasmides dérivent des travaux de Jim Boulanger. Après sélection des transformants, les levures sont transférées sur gélose 5'FOA. Aucun des plasmides testés ne permet une bonne croissance sur ce milieu. On voit dans certains cas apparaître quelques micro-colonies, mais cette croissance est très faible comparativement à celle obtenue avec la souche BY4705 sauvage. Donc, dans les conditions utilisées, aucun gène ou fragment de chromosome testé ne complémente clairement la perte du plasmide contenant RoBP1 chez la levure dépendante de RoBP1.

Les expériences de complémentation effectuées par Jim Boulanger suggéraient par contre que la dépendance à RoBP1 provenait de mutations dans 2 gènes différents. Il avait identifié le gène CAF4 présent sur le chromosome XI comme étant un de ces gènes. Le gène UBC5 sur le chromosome IV était l'autre candidat. S'il s'agit effectivement des 2 homologues fonctionnels de RoBP1, la double délétion devrait être létale pour la levure.

4.6 Délétion des gènes UBC5 et CAF4 dans la levure BY4705

Un allèle du gène UBC5 est remplacé par le marqueur LEU2 dans la souche

BY4705 diploïde. Après sélection des clones positifs sur milieu restrictif Yc-Leu, la délétion est confirmée par PCR en utilisant les amorces For-UBC5 et Rev-UBC5 (voir figure 5A). La taille des fragments obtenus concorde avec les tailles attendues : la bande de 1,0 kb correspond à l'allèle UBC5 sauvage et la bande de 1,6 kb correspond à l'allèle remplacé par LEU2 (Figure 13). L'ADN de la souche BY4705 sauvage produit uniquement la bande de 1,0 kb.

De la même façon, la souche BY4705 diploïde est transformée avec un fragment d'ADN contenant le marqueur TRP1 afin de remplacer un allèle du gène CAF4. La délétion est confirmée par PCR en utilisant les amorces CAF4-3For et CAF4-2Rev (voir figure 5B). L'ADN de la souche BY4705 sauvage produit uniquement une bande de 2,4 kb correspondant à l'allèle CAF4 sauvage (Figure 14). L'ADN de la souche de délétion produit une bande de 1,5 kb correspondant à l'allèle remplacé par TRP1 en plus de la bande correspondant à l'allèle sauvage.

Un allèle du gène CAF4 est remplacé par le marqueur HIS3 dans la souche diploïde hétérozygote pour la délétion du gène UBC5. La délétion est également confirmée par PCR en utilisant les amorces CAF4-3For et CAF4-2Rev (voir figure 6). Malgré la migration décalée du marqueur de poids moléculaire, on peut voir que la taille des fragments obtenus concorde avec les tailles attendues : la bande de 2,4 kb correspond à l'allèle CAF4 sauvage et la bande de 1,5 kb correspond à l'allèle remplacé par HIS3 (Figure 15). L'ADN de la souche BY4705 sauvage produit uniquement la bande de 2,4 kb.

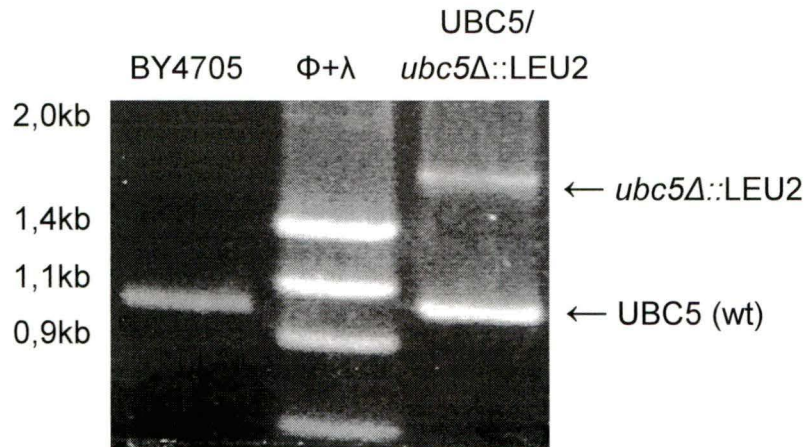


Figure 13 : Vérification du remplacement du gène UBC5 par LEU2

Réaction de PCR utilisant l'ADN génomique de la souche BY4705 diploïde dans laquelle un allèle du gène UBC5 est remplacé par le marqueur LEU2 et l'ADN génomique de la souche BY4705 diploïde sauvage. Les amorces utilisées sont For-UBC5 et Rev-UBC5. Les produits de PCR sont analysés sur gel d'agarose 1%. Le marqueur de poids moléculaire (Φ + λ) est un mélange des marqueurs λ DNA / *HindIII* et Φ X174 DNA / *HaeIII*. Les chiffres à gauche correspondent à la taille des bandes du marqueur.

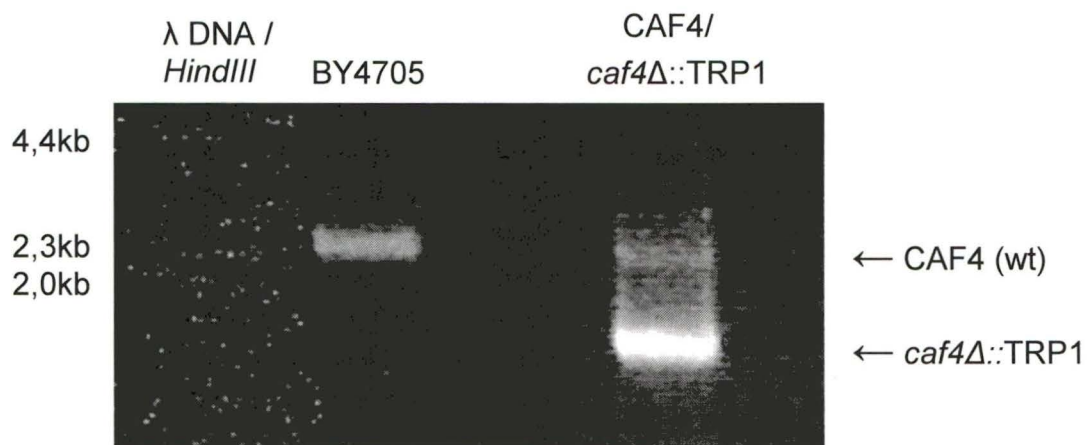


Figure 14 : Vérification du remplacement du gène CAF4 par TRP1

Réaction de PCR utilisant l'ADN génomique de la souche BY4705 diploïde dans laquelle un allèle du gène CAF4 est remplacé par le marqueur TRP1 et l'ADN génomique de la souche BY4705 diploïde sauvage. Les amorces utilisées sont CAF4-3For et CAF4-2Rev. Les produits de PCR sont analysés sur gel d'agarose 1%. Le marqueur de poids moléculaire est λ DNA / *HindIII*. Les chiffres à gauche correspondent à la taille des bandes du marqueur.

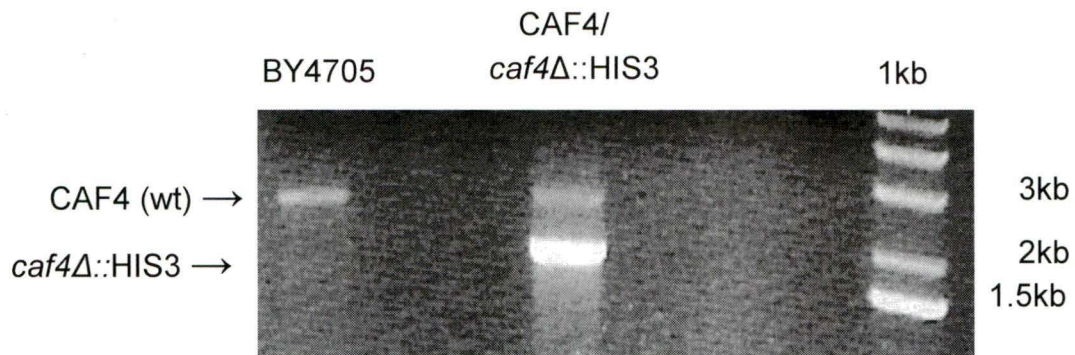


Figure 15 : Vérification du remplacement du gène CAF4 par HIS3

Réaction de PCR utilisant l'ADN génomique de la souche BY4705 diploïde sauvage et l'ADN génomique de la souche BY4705 diploïde dans laquelle un allèle de chaque gène CAF4 et UBC5 est remplacé par les marqueurs HIS3 et LEU2, respectivement. Les amorces utilisées sont CAF4-3For et CAF4-2Rev. Les produits de PCR sont analysés sur gel d'agarose 1%. Les chiffres à droite correspondent à la taille des bandes du marqueur 1kb.

La sporulation et microdissection de 10 tétrades de la souche diploïde hétérozygote pour la délétion du gène UBC5 montre une ségrégation des spores de 2 : 2 (vie : mort) sur milieu YEPD. Par contre, cette ségrégation ne concorde pas nécessairement avec la ségrégation du marqueur LEU2, vérifiée par ensemencement des spores sur milieu Yc-Leu. En effet, certaines spores peuvent croître sur les milieux YEPD et Yc-Leu indiquant que le gène UBC5 n'est pas essentiel (Tableau 4). Le même résultat est obtenu avec la souche diploïde hétérozygote pour la délétion du gène CAF4. Donc, malgré une ségrégation 2 : 2, les gènes UBC5 et CAF4 ne sont pas des gènes essentiels. Notons cependant que les souches diploïdes contiennent quand même une mutation létale dont l'identité est inconnue.

De la même façon, la sporulation et microdissection de 30 tétrades de la souche diploïde hétérozygote pour la délétion des gènes UBC5 et CAF4 montrent une ségrégation 2 : 2 (vie : mort) qui ne concorde pas avec la ségrégation des marqueurs LEU2 et HIS3. Ainsi, plusieurs spores peuvent croître sur le milieu restrictif Yc-Leu-His. Donc, les gènes CAF4 et UBC5 ne sont pas synthétiquement létaux et ne sont pas les homologues fonctionnels de RoBP1.

		Croissance sur milieu	
Tétrade	Spore	YEPD	Yc-Leu
1	A	+	-
	B	-	ND
	C	-	ND
	D	+	+
2	A	+	+
	B	+	-
	C	-	ND
	D	-	ND
3	A	-	ND
	B	+	+
	C	+	+
	D	-	ND
4	A	+	-
	B	-	ND
	C	+	+
	D	-	ND

Tableau 4 : Exemple de ségrégation des spores issues de la microdissection de la souche diploïde contenant la délétion du gène UBC5.

(+) Croissance de la spore sur le milieu. (-) Absence de croissance sur le milieu.
 (ND) Non déterminé. Les cases ombragées indiquent les spores viables ayant remplacé le gène UBC5 par le marqueur LEU2.

Pseudogènes hY

4.7 Identification de 966 pseudogènes des ARNs hY

La recherche de séquences homologues aux ARNs hY, et à différents petits ARNs non-codants, dans le génome humain avec l'outil MegaBLAST a révélé une grande quantité de pseudogènes potentiels qui remplissent les critères de sélection mentionnés précédemment (Tableau 5). Les ARN de transfert pour l'arginine et la lysine présentent respectivement 5 et 35 séquences homologues et l'ARN ribosomal 5S en présente 659. La recherche a identifié un total de 1366 séquences homologues aux petits ARNs nucléaires U1-6, dont 1085 (79%) provenant uniquement de l'ARN U6. Ces nombres sont similaires aux nombres obtenus du séquençage du génome humain (LANDER et al., 2001). Cette recherche a également identifié dans le génome un total de 966 séquences homologues aux ARNs hY, avec une faible probabilité d'obtenir ces séquences par chance (E-value <0,001 pour 99% des séquences). Les séquences dérivées des ARNs hY3 et hY1 sont les plus abondantes avec 442 copies (46%) et 368 copies (38%), respectivement. Les séquences homologues aux ARNs hY4 et hY5 sont beaucoup moins nombreuses, avec 148 copies (15%) et 8 copies (<1%), respectivement. Ces résultats sont aussi en accord avec les données obtenues du séquençage du génome humain (LANDER et al., 2001). Une analyse attentive de la région en amont des séquences homologues aux ARNs hY n'a pas permis l'identification de promoteurs connus, indiquant que ces séquences sont vraisemblablement des pseudogènes.

Pseudogènes	Nombre
hY1	368
hY3	442
hY4	148
hY5	8
U1	91
U2	46
U3	45
U4	70
U5	29
U6	1085
ARNt ^{Arg}	5
ARNt ^{Lys}	35
ARNr 5S	659

Tableau 5 : Nombre de séquences homologues à différents petits ARNs non-codants dans le génome humain.

Les résultats sont obtenus avec l'outil MegaBLAST et incluent seulement les séquences ayant plus de 75% d'identité sur au moins 50% de la séquence du gène correspondant. Ces nombres excluent les séquences 100% identiques au gène correspondant.

Les pseudogènes hY sont distribués sur tous les chromosomes humains, tandis que les gènes codant pour les 4 ARNs hY sont groupés à une extrémité du chromosome 7 (Figure 16). De façon générale, le nombre de pseudogènes hY par chromosome est proportionnel à la taille du chromosome, avec un maximum de 94 copies sur le chromosome 1 (Figure 17). Par contre, les chromosomes 1, 7, 12 et 17 montrent une densité relativement plus élevée que la moyenne en pseudogènes hY, tandis que les chromosomes 8 et Y sont moins densément peuplés (Tableau 6). Les pseudogènes hY sont notamment peu abondants sur le chromosome Y, avec seulement une copie. Cette distribution est très similaire à celle observée pour les éléments Alu (GROVER et al., 2004) et les rétropseudogènes des protéines ribosomales (ZHANG et al., 2002), avec les exceptions d'une plus grande densité de pseudogènes hY sur le chromosome 7 et une plus faible densité sur le chromosome 19.

4.8 Conservation et âge des pseudogènes hY

Parmi les 966 pseudogènes hY identifiés, aucun n'est 100% identique aux séquences des gènes hY fonctionnels correspondants. Les différences de séquence entre les pseudogènes hY et les gènes hY sont dispersées sur toute leur longueur. Pour les pseudogènes hY, le taux moyen de mutation pour une position donnée, par rapport aux séquences des gènes hY fonctionnels correspondants, est de $10 \pm 4\%$. Par contre, il existe 3 déviations significatives de ce patron de mutation apparemment aléatoire. La première consiste en une mutation quasi constante à au moins une

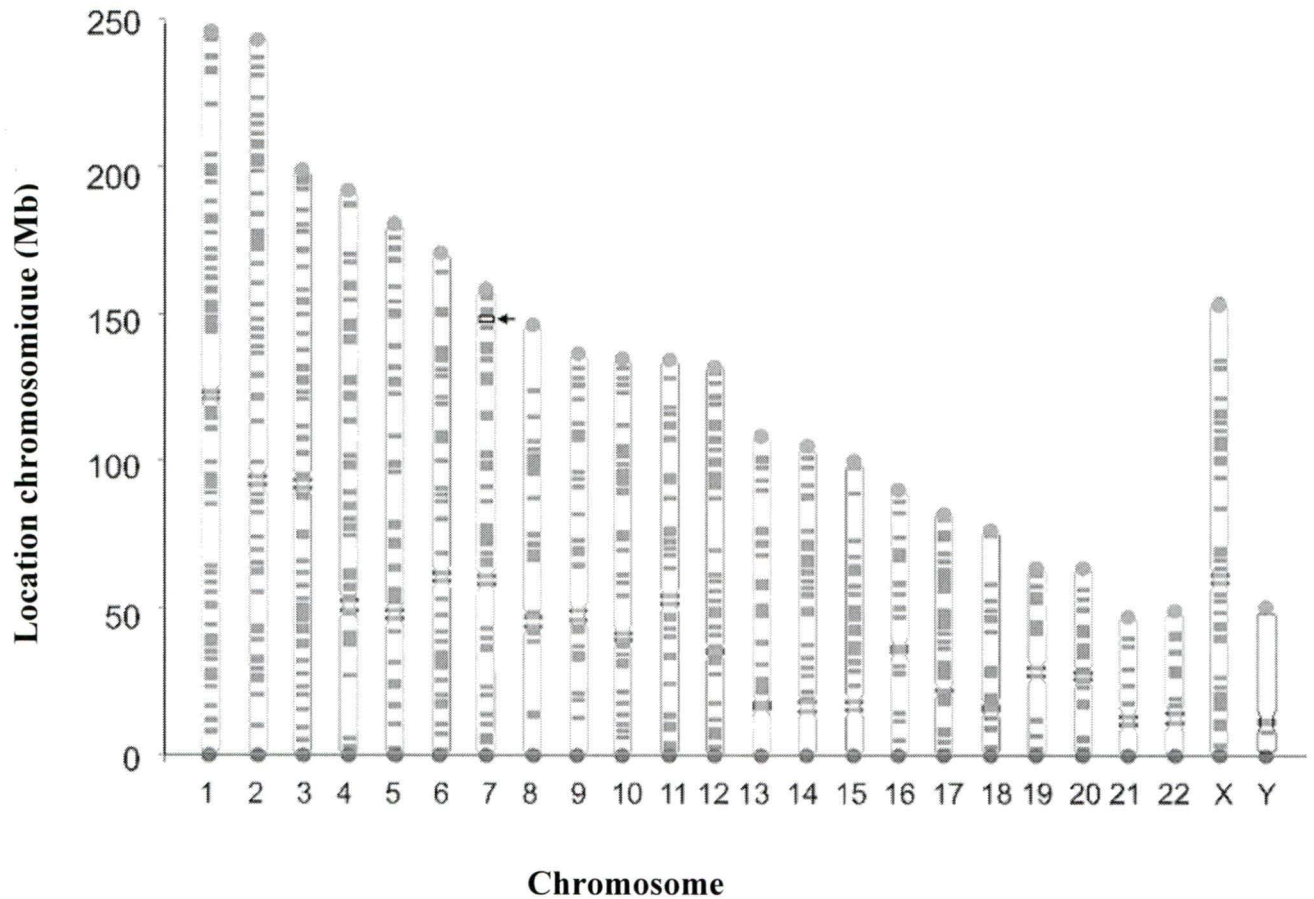


Figure 16 : Distribution des pseudogènes hY dans le génome humain.

Les 24 chromosomes humains sont montrés verticalement de la gauche à la droite. Les pseudogènes sont représentés par de courts traits horizontaux, les télomères par des points et les centromères par une paire de longs traits horizontaux. La position des gènes hY fonctionnels est indiquée par la flèche et la petite boîte à l'extrémité du chromosome 7. (Mb) Mégabase d'ADN.

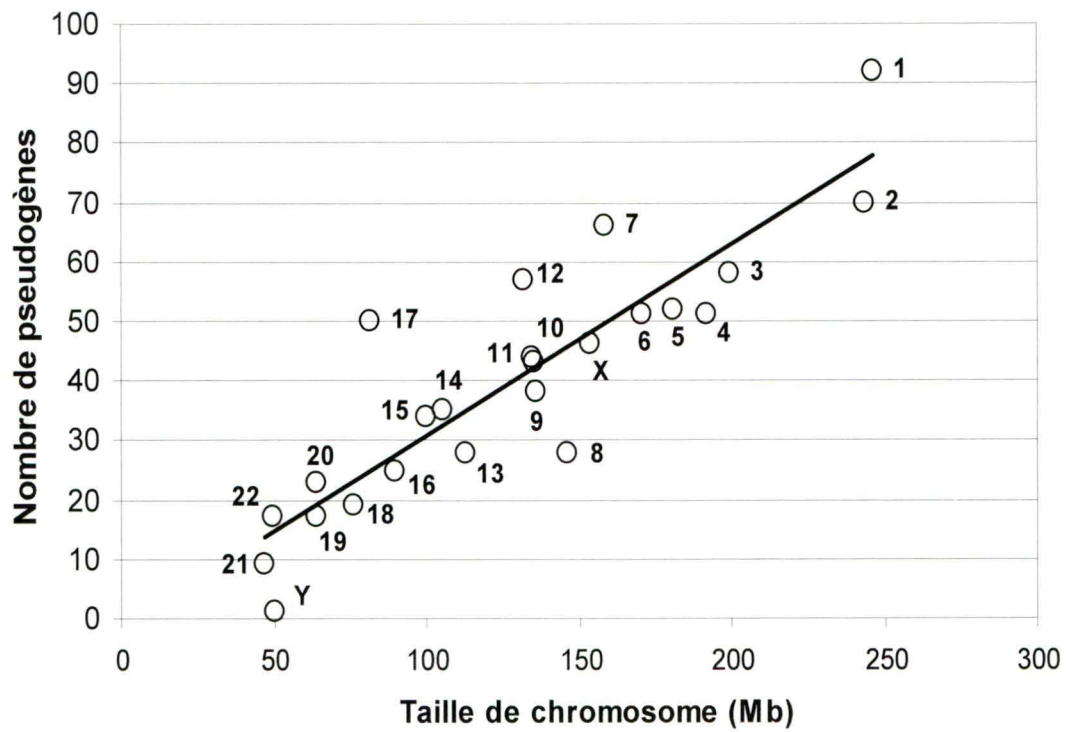


Figure 17 : Corrélation entre le nombre de pseudogènes hY et la taille des chromosomes.

Les cercles numérotés représentent les chromosomes.

Chromosome	Nombre de pseudogènes	Taille du chromosome (Mb)	Pseudogènes/Mb
1	94	246	0,38
2	72	244	0,30
3	62	199	0,31
4	52	192	0,27
5	53	181	0,29
6	53	171	0,31
7	63	159	0,40
8	29	146	0,20
9	37	136	0,27
10	42	135	0,31
11	47	134	0,35
12	56	132	0,42
13	28	113	0,25
14	38	105	0,36
15	35	100	0,35
16	25	90	0,28
17	50	82	0,61
18	20	76	0,26
19	17	64	0,27
20	23	64	0,36
21	6	47	0,13
22	16	49	0,32
X	47	154	0,31
Y	1	50	0,02
Total	966	3070	0,31

Tableau 6 : Densité des pseudogènes hY dans le génome humain

Nombre de pseudogènes hY par mégabase (Mb) d'ADN.

position du dinucléotide CG aux positions 9 et 10 (Figure 18). Au moins un de ces 2 nucléotides est muté dans plus de 80% des pseudogènes correspondant aux ARNs hY1, hY3 et hY4. Les positions 9 et 10 correspondent à la région reconnue comme étant essentielle à la liaison de la protéine Ro60 aux ARNs Y (GREEN et al., 1998). La seconde déviation du bruit de fond est un taux de mutation de 80% dans les nucléotides formant la région riche en pyrimidines située au milieu de l'ARN hY1 (Figure 18). Cette région est impliquée dans la liaison des protéines hnRNP K et PTB à l'ARN hY1. Finalement, 80, 40 et 25% des pseudogènes sont dépourvus respectivement de 1, 2 et 3 uraciles à leur extrémité 3' (Figure 18). Cette extrémité poly(U) en 3' est importante pour la liaison de la protéine La. L'observation de ces trois déviations suggère que des mutations à des positions spécifiques, qui limitent l'association avec certaines protéines, pourraient contribuer à augmenter l'efficacité de rétrotransposition des ARNs hY. Mais il se pourrait aussi qu'une trop grande quantité de pseudogènes fonctionnels soit nocive pour la cellule. Les pseudogènes mutés seraient ainsi retenus de façon préférentielle.

Le pourcentage de divergence des pseudogènes hY par rapport aux séquences des gènes hY fonctionnels suit une distribution normale, avec un maximum à 9%. En comparaison, la distribution de la divergence pour les séquences Alu présente un maximum à 7% (Figure 19A; LANDER et al., 2001). Il a été estimé qu'en absence de pression de sélection, chaque nucléotide a un taux de mutation de 1% à chaque 6,7 million d'années (KUMAR et al., 2001). En apparence, cette différence de distribution de la divergence suggère que la plupart des pseudogènes hY seraient apparus avant les éléments Alu. Cependant, considérant que certaines positions dans

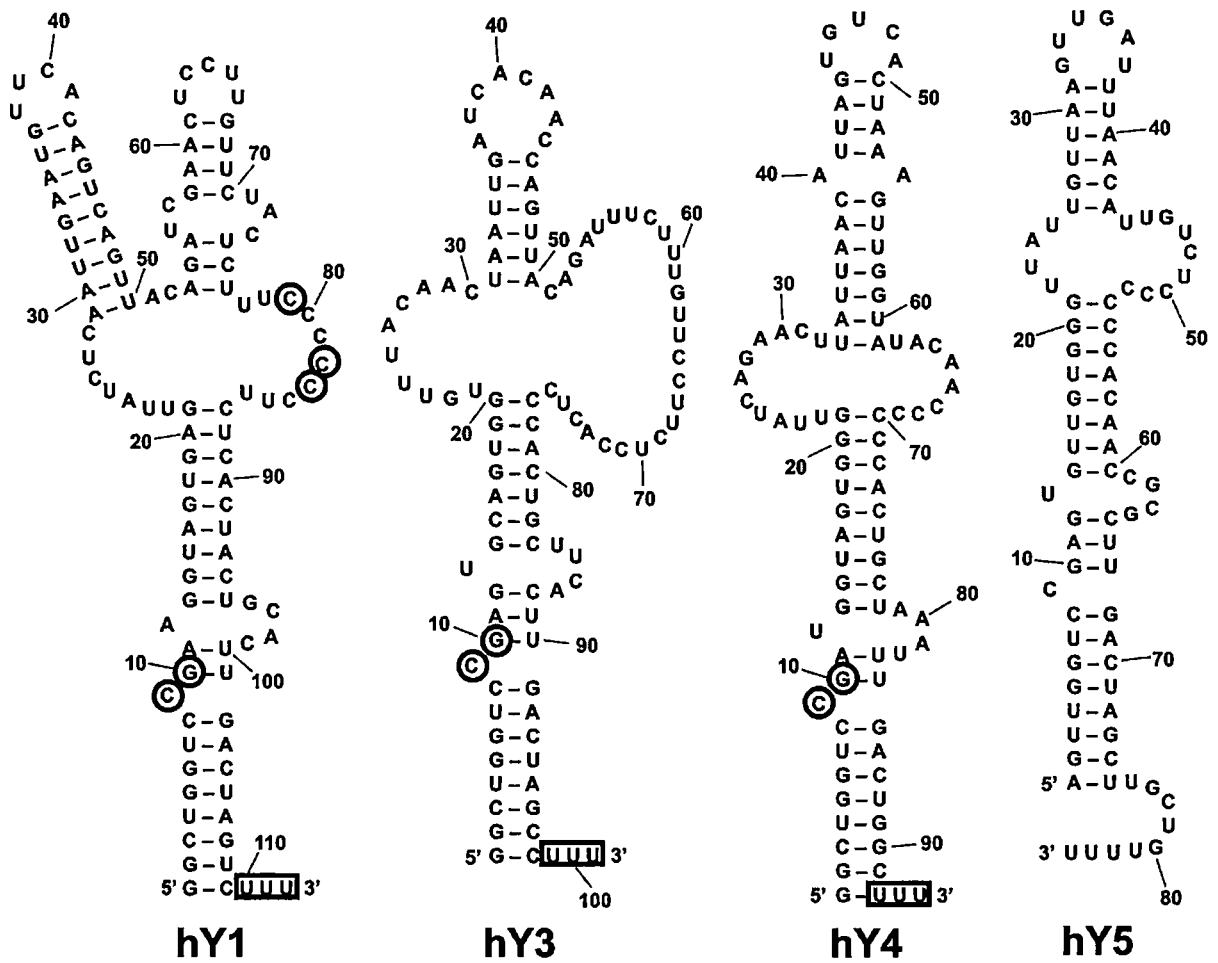


Figure 18: Mutations les plus fréquentes dans les pseudogènes hY.

Les nucléotides encerclés correspondent aux positions les plus fréquemment mutées dans les pseudogènes hY. Les nucléotides dans des boîtes correspondent aux nucléotides les plus souvent absents des pseudogènes hY.

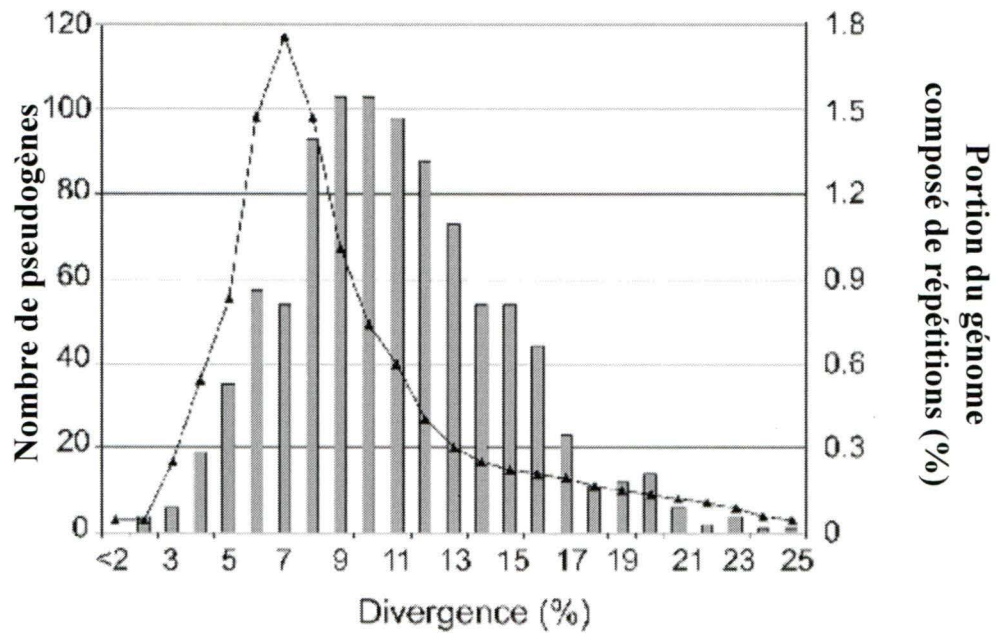
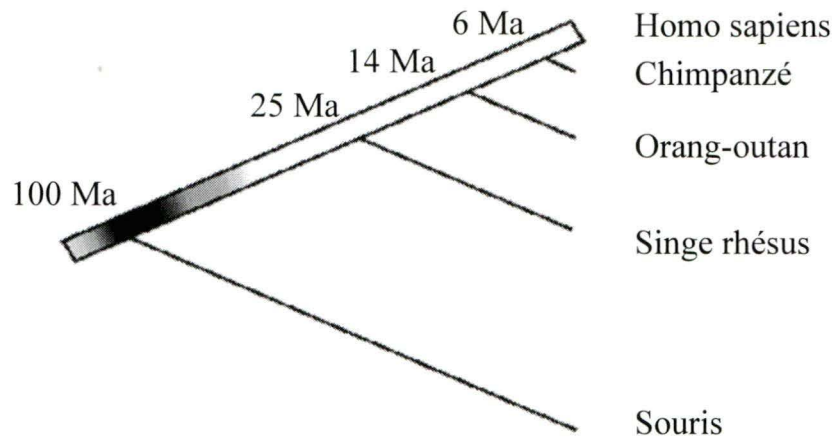
A**B**

Figure 19 : Conservation et âge des pseudogènes hY.

(A) Distribution des pseudogènes selon leur pourcentage de divergence. La distribution des pseudogènes hY est illustrée par l'histogramme et utilise l'échelle de gauche. La distribution des éléments Alu est illustrée par la courbe avec des triangles et utilise l'échelle de droite. (B) Arbre phylogénétique de quelques primates et de la souris (adapté de GOODMAN et al., 1998). L'intensité de la couleur de la ligne indique grossièrement la période la plus active en apparition de pseudogènes hY.

les pseudogènes hY sont presque uniformément mutées, et que ces mutations pourraient avoir été présentes au niveau de l'ARN avant l'intégration dans l'ADN, les éléments Alu et les pseudogènes hY pourraient être apparus à la même période. Comme la plupart des pseudogènes hY montrent une divergence de moins de 15% par rapport aux séquences des gènes hY fonctionnels correspondants, une grande proportion des pseudogènes sont apparus il y a moins de 100 millions d'années, soit après la radiation des primates et des rongeurs (Figure 19B). Cette hypothèse est confirmée par une recherche de séquences homologues aux ARNs mY1 dans le génome de la souris en utilisant le programme BLAT. Seulement 24 pseudogènes mY3 et 3 pseudogènes mY1 sont identifiés (Tableau 7). De plus, aucun de ces pseudogènes n'est présent aux positions génomiques correspondantes chez l'humain. Donc, l'apparition des pseudogènes hY est survenue après la divergence des rongeurs et des primates.

Au contraire, l'humain et le chimpanzé, dont la divergence remonte à 6 millions d'années, devraient avoir la plupart des pseudogènes Y en commun. Les séquences des ARNs Y1, Y3 et Y4 humains sont identiques à ceux du chimpanzé, mais la séquence de l'ARN Y5 présente une insertion d'un nucléotide dans la version du chimpanzé. La distribution des pseudogènes Y trouvés chez le chimpanzé à l'aide du programme BLAT est presque identique à celle de l'humain, mais le nombre de pseudogènes Y trouvés chez le chimpanzé est légèrement inférieur, soit environ 90% du nombre trouvé chez l'humain (Tableau 7). Cette différence suggère que de nouveaux pseudogènes sont apparus chez l'humain ou ont été perdus chez le chimpanzé depuis les derniers 6 millions d'années d'évolution. Alternativement,

	Humain	Chimpanzé	Souris
Y1	1,00	0,89	0,06
Y3	1,02	0,96	0,01
Y4	0,60	0,52	NA
Y5	0,03	0,02	NA

Tableau 7 : Nombre relatif de pseudogènes Y chez différentes espèces.

Les nombres sont relatifs au nombre de pseudogènes Y1 chez l'humain, selon les résultats obtenus avec l'outil BLAT. Il n'y a pas de Y4 et Y5 chez la souris (NA).

certaines des pseudogènes manquants chez le chimpanzé pourraient être situés dans des régions du génome qui ne sont pas encore complètement assemblées.

4.9 Localisation génomique des pseudogènes hY et statistiques

La position génomique des séquences homologues aux ARNs hY par rapport aux gènes codants pour des protéines a été déterminée (Tableau 8). Parmi les 966 pseudogènes hY trouvés dans le génome humain, 403 (42%) sont localisés dans des gènes codants pour des protéines. Par contre, seulement 3 de ces pseudogènes se trouvent dans des exons, plus précisément dans des régions non-traduites. Les trois gènes sont ALDH9A1, CCL19 et MPRG (aldéhyde déshydrogénase, petit précurseur inductible de la cytokine A19 et récepteur γ membranaire de la progestine). Aucune fonction ou structure commune entre ces gènes n'est apparente actuellement.

Sachant qu'environ 25% du génome est composé de séquences géniques (JASINSKA & KRZYZOSIAK, 2004), on peut calculer la densité des pseudogènes hY situés dans les régions intergéniques et intragéniques. On retrouve 0,53 pseudogène hY par 10^6 bases d'ADN pour les régions intragéniques comparativement à 0,24 pseudogène hY par 10^6 bases d'ADN pour les régions intergéniques (tableau 8). Il semble donc y avoir un biais pour la localisation des pseudogènes hY dans des gènes.

Il n'y a aucune différence significative entre les pseudogènes hY situés dans les introns et ceux situés dans les régions intergéniques tant au niveau de la conservation et de la taille de leur séquence, qu'au niveau des mutations spécifiques.

Lorsque leur orientation est comparée à l'orientation des gènes dans lesquels ils se trouvent, les pseudogènes hY introniques sont presque autant orientés dans le même sens (52%) que dans le sens opposé (48%; Tableau 8). Encore une fois, aucune différence significative n'est observée entre les pseudogènes hY dans l'orientation sens et ceux dans l'orientation antisens.

Les pseudogènes dérivés des 4 ARNs hY sont en moyenne près de 90% identiques aux gènes fonctionnels correspondants. La taille des pseudogènes hY est en moyenne environ 10% plus courte que la taille des gènes fonctionnels correspondants (Tableau 8).

Le contenu moyen en GC des pseudogènes hY (42%) est légèrement inférieur au contenu moyen des gènes hY fonctionnels (44,5%; Tableau 8). Ce résultat peut s'expliquer par le taux relativement élevé de déamination spontanée transformant les C en T dans l'ADN. Les régions de 5 kb flanquant les pseudogènes hY montrent un contenu en GC moyen de 41% (Tableau 8), et ce pourcentage ne varie pas significativement selon la position intronique ou intergénique des pseudogènes. Ce contenu en GC est comparable à celui observé pour les pseudogènes des protéines ribosomales (ZHANG et al., 2002), et se situe entre l'environnement riche en GC des éléments Alu et l'environnement riche en AT des éléments L1 (LANDER et al., 2001).

4.10 Rétrotransposition des ARN hY par la machinerie L1

Une grande proportion des répétitions dispersées retrouvées dans le génome

	Position des pseudogènes							Contenu en GC			
	Intragénique ^a			Intergénique ^d	Pseudogènes/Mb ^e		Identité moyenne ^f				
	Sens ^b	Antisens ^c	Total		Intragénique	Intergénique		Gènes hY	Pseudogènes hY	Régions flanquantes ^h	
hY1	80	82	162	206	0,21	0,09	87,9%	91,5%	44,6%	42,1%	42,5%
hY3	103	71	174	268	0,23	0,12	90,6%	90,3%	45,5%	42,6%	40,0%
hY4	24	38	62	86	0,08	0,04	90,7%	87,2%	42,6%	40,5%	40,6%
hY5	4	1	5	3	0,007	0,001	89,4%	86,3%	45,2%	42,9%	40,6%
Total	211	192	403	563	0,53	0,24	89,6%	88,8%	44,5%	42,0%	41,0%

Tableau 8 : Statistiques globales des pseudogènes hY.

^aNombre de pseudogènes hY situés dans un gène.

^bNombre de pseudogènes hY situés dans un gène, dans la même orientation que ce gène.

^cNombre de pseudogènes hY situés dans un gène, dans l'orientation opposée à ce gène.

^dNombre de pseudogènes hY situés dans une région intergénique.

^eNombre de pseudogènes par mégabase (Mb) d'ADN

^fIdentité moyenne de la séquence des pseudogènes hY comparée aux séquences des gènes hY correspondants.

^gTaille moyenne des pseudogènes hY divisée par la taille des gènes hY correspondants.

^hContenu en GC de la région de 5 kb flanquant les pseudogènes hY.

humain sont des rétrotransposons ou ont été générées par des rétrotransposons. Tout comme les éléments Alu, les ARNs hY ne codent pour aucune protéine, leur répartition dans le génome dépendant forcément de l'action de protéines encodées par d'autres éléments. Le principal responsable de l'expansion de ces séquences répétées dans le génome humain est l'élément L1. Les signatures caractéristiques d'un événement de rétrotransposition par la machinerie L1 consistent en une queue poly(A) de taille variable à l'extrémité 3' de l'élément rétrotransposé et des duplications du site d'insertion (TSDs) d'environ 15 nucléotides flanquant l'élément. De plus, le TSD à l'extrémité 5' débute habituellement par une séquence similaire à TTAAAA, le site consensus de clivage de l'endonuclease L1 (Figure 20A; OSTERTAG & KAZAZIAN, 2001).

Le programme RTAnalyser, conçu par Jean-François Lucier, a été utilisé pour analyser les séquences des pseudogènes hY à la recherche de signatures de rétrotransposition par L1. Ce programme identifie d'abord les extrémités 5' et 3' des pseudogènes hY et cherche ensuite pour une queue poly(A) en 3' et des TSDs de part et d'autre du pseudogène. Pour évaluer la qualité de ces signatures, un système de pointage (RetroScore) a été établi considérant plusieurs paramètres, incluant la taille de la queue poly(A) et des TSDs, leur position, l'homologie entre les TSDs, etc (voir Matériel et Méthodes). Les pseudogènes obtenant un RetroScore de plus de 40 sont considérés comme ayant une haute probabilité d'avoir été générés par rétrotransposition. Certaines séquences ont été utilisées comme contrôles négatifs afin de confirmer qu'un RetroScore de 40 est spécifique pour l'identification de signature L1. Un fragment intronique du gène Ro60 d'une taille de 120 nt et les

séquences de 2 ARNs de transfert (ARNt^{arg} et ARNt^{lys}) ont obtenu des pointages largement en dessous du seuil de 40, avec l'exception de quelques pseudogènes d'ARNt montrant des signatures de rétrotransposition évidentes et donnant des pointages bien supérieurs à 40. De la même façon, les gènes fonctionnels codant pour les ARNs hY1, hY3, hY4 et hY5 donnent des pointages de 0, 10, 15 et 22, respectivement.

La figure 20B montre un exemple typique d'un pseudogène hY1 avec d'excellentes signatures de rétrotransposition. La queue poly(A), d'une taille de 14 nt et composée de 79% de A, est située immédiatement en 3' de l'extrémité du pseudogène. Les TSDs sont d'une taille de 17 nt avec seulement un mésappariement et sont situés immédiatement en 5' du pseudogène (TSD 5') et en 3' de la queue poly(A) (TSD 3'). De plus, le site consensus de clivage de l'endonuclease L1 (TTAAAA) est correctement positionné, chevauchant le TSD en 5' de 4 nt. Parmi tous les pseudogènes hY identifiés, 63% sont flanqués par de telles signatures de rétrotransposition par L1 (Tableau 9).

L'examen détaillé des séquences des pseudogènes hY obtenant un RetroScore sous le seuil de 40 révèle que, dans plusieurs cas, un des TSDs (habituellement le TSD 3') a été détruit ou déplacé par l'intégration indépendante d'un élément Alu (Figure 20C). Lorsque cet élément Alu est retiré, le TSD qui était scindé en 2 peut être à nouveau associé à son homologue intact situé à l'autre extrémité du pseudogène. Cette observation suggère que dans ces cas, les pseudogènes hY étaient déjà présents lors de l'insertion de l'élément Alu. L'hypothèse est confirmée dans certains cas où les TSDs des pseudogènes correspondant dans le génome du

	RetroScore au-dessus du seuil	
	Pseudogènes hY	Pseudogènes hY-Alu
hY1	58%	67%
hY3	64%	69%
hY4	71%	74%
hY5	63%	75%
Total	63%	69%

Tableau 9 : Pseudogènes hY montrant des signatures de rétrotransposition par L1.

Pourcentage de pseudogènes hY obtenant un RetroScore supérieur au seuil de 40, considéré comme étant spécifique pour l'identification de signatures de rétrotransposition par L1. Le RetroScore est calculé par le programme RTAnalyser en tenant compte de plusieurs paramètres (voir matériel et méthodes). La dernière colonne montre le pourcentage de pseudogènes hY obtenant un RetroScore supérieur au seuil de 40 en considérant l'impact de l'insertion d'éléments Alu (Pseudogènes hY-Alu).

chimpanzé ne sont pas détruits/déplacés par un élément Alu (Figure 20C). Cette fréquence d'insertion d'éléments Alu près des pseudogènes hY peut s'expliquer par le fait que le TSD et la queue poly(A) adjacente forment des séquences similaires au site consensus de coupure par l'endonuclease L1. Lorsque l'impact de l'insertion des éléments Alu est pris en considération pour l'évaluation du RetroScore, la proportion des pseudogènes hY montrant des signatures de rétrotransposition par L1 convaincantes passe à près de 70% (Tableau 9). Donc, la grande majorité des pseudogènes hY pourraient être reconnue comme provenant d'un événement de rétrotransposition par la machinerie L1.

De plus, jusqu'à 12% des pseudogènes hY sont tronqués à leur extrémité 5' d'au moins 10% de leur séquence lorsque comparés aux gènes fonctionnels (Figure 20D). Comme la grande majorité des éléments L1 trouvés dans le génome humain sont tronqués à leur extrémité 5', vraisemblablement à cause d'un arrêt prématuré de la transcriptase inverse, cette observation est une autre évidence que les pseudogènes hY ont été générés par la machinerie de rétrotransposition de L1. La transcriptase inverse de L1, bien que peu processive, a probablement moins de difficulté à compléter la transcription inverse d'un ARN hY (85-112 nt) que celle d'un ARN L1 (6 kb), ce qui expliquerait que seulement 12% des pseudogènes hY soient tronqués.

La machinerie de rétrotransposition de L1 peut également participer à la création de pseudogènes chimériques, formés de l'association d'un élément L1 ou Alu (3') avec un snARN (5'). La présence de plus de 300 petits ARNs non-codants (ARNt, ARNr, snARN, snoARN, micro ARN) a été recherchée à proximité des pseudogènes hY, mais seulement des séquences Alu ont été identifiées. Des éléments

Alu se retrouvent à proximité (<300 nt) de plus de la moitié des pseudogènes hY. Cependant, une analyse détaillée des séquences révèle que moins de 5% des pseudogènes hY pourraient correspondre à des pseudogènes chimériques authentiques.

4.11 Pseudogènes hY dupliqués

Bien que la majorité des pseudogènes hY semblent provenir d'événements de rétrotransposition indépendants, environ 5% représentent des duplications de segments d'ADN contenant déjà des pseudogènes hY. Ce résultat est obtenu en analysant l'alignement des régions adjacentes aux pseudogènes hY. Par exemple, 15 membres de la famille de gènes PMS2 (Postmeiotic segregation increased 2; NICOLAIDES et al., 1995) contiennent un pseudogène hY3 très similaire dans la même région du même intron et dans la même orientation. PMS2 est un membre de la machinerie de réparation des mésappariements dans l'ADN (SHIN-DARLAK et al., 2005). Vraisemblablement, un événement initial de rétrotransposition d'un ARN hY3 est survenu dans l'ancêtre du gène PMS2. Par la suite, la famille de gènes PMS2 a été créée suite à une succession d'événements de duplication de ce gène ancestral. Pendant ce temps, des mutations ponctuelles se sont accumulées dans les pseudogènes hY3 présents chez les membres de la famille PMS2. Les différences de séquence entre ces pseudogènes dupliqués peuvent être utilisées pour établir la séquence d'apparition de ces gènes dupliqués. La recherche a identifié 15 autres groupes

similaires contenant de 2 à 7 pseudogènes hY, mais ces duplications sont survenues dans des régions intergéniques.

hY et épissage

4.12 Modèle d'épissage alternatif hnRNP A1

Étant donné la répartition préférentielle des pseudogènes hY dans des introns, il est possible que ces séquences aient un impact sur l'épissage. Pour vérifier si la présence de séquences introniques correspondant aux ARNs hY a un effet sur la modulation de l'épissage alternatif, 2 modèles bien étudiés dans le laboratoire du Dr Benoit Chabot ont été utilisés. Le premier modèle est dérivé du gène codant pour la protéine hnRNP A1. Ce modèle avait été développé pour étudier l'épissage de l'ARN pré-messager du gène hnRNP A1, plus spécifiquement au niveau de la régulation de l'épissage de l'exon alternatif 7B. Il est constitué d'une partie des séquences des exons 7 et 7B, et de leur site d'épissage 5' respectifs, ainsi que d'un site d'épissage 3' dérivé de l'exon L2 de l'adénovirus. Dans ce modèle, 2 sites d'épissage 5' sont en compétition avec 1 site d'épissage 3' (Figure 7). Cette unité d'épissage alternatif peut être transcrite et l'ARN pré-messager épissé en présence d'un extrait nucléaire de cellules. Pour vérifier l'influence sur l'épissage de séquences régulatrices potentielles, il est possible de positionner ces séquences en aval ou en amont de l'exon alternatif 7B et d'observer la modulation du choix du site d'épissage 5'.

4.12.1 Les séquences hY3 introniques modulent l'épissage *in vitro*

4.12.1.1 Séquences hY3 en aval de l'exon alternatif

L'épissage de l'ARN pré-messager transcrit à partir de l'unité d'épissage alternatif hnRNP A1 donne essentiellement 3 produits lorsque analysé par RT-PCR (Figure 21A). Comme la réaction d'épissage *in vitro* n'est pas une réaction extrêmement efficace, l'ARN pré-messager n'est jamais complètement épissé. Le produit dont la migration dans le gel est la plus lente, qui est le produit de plus grande taille, représente l'ARN pré-messager non-épissé. Le produit de taille intermédiaire représente le produit d'épissage résultant de la sélection du site d'épissage 5' de l'exon 7B et est nommé produit proximal. Le produit ayant la plus petite taille représente le produit distal, soit le produit issu de l'utilisation du site d'épissage 5' de l'exon 7.

L'ARN pré-messager de base, transcrit à partir du plasmide 45, est épissé majoritairement au site d'épissage 5' de l'exon 7B. Le site d'épissage 5' proximal est ainsi sélectionné dans 79% des cas, tandis que le site d'épissage distal n'est utilisé que dans 21% des cas (Figure 21). Par contre, la présence de sites de liaisons de haute affinité pour la protéine hnRNP A1 de part et d'autre de l'exon alternatif 7B (plasmide 36) déplace le choix du site d'épissage 5' proximal vers le site 5' distal. Dans ce cas, le site d'épissage 5' proximal n'est plus utilisé que dans 17% des cas. Le plasmide 45hY3 a été construit en insérant la séquence d'ADN correspondant à l'ARN hY3, au niveau du site de restriction *EcoRV*, à 30 nucléotides en aval de

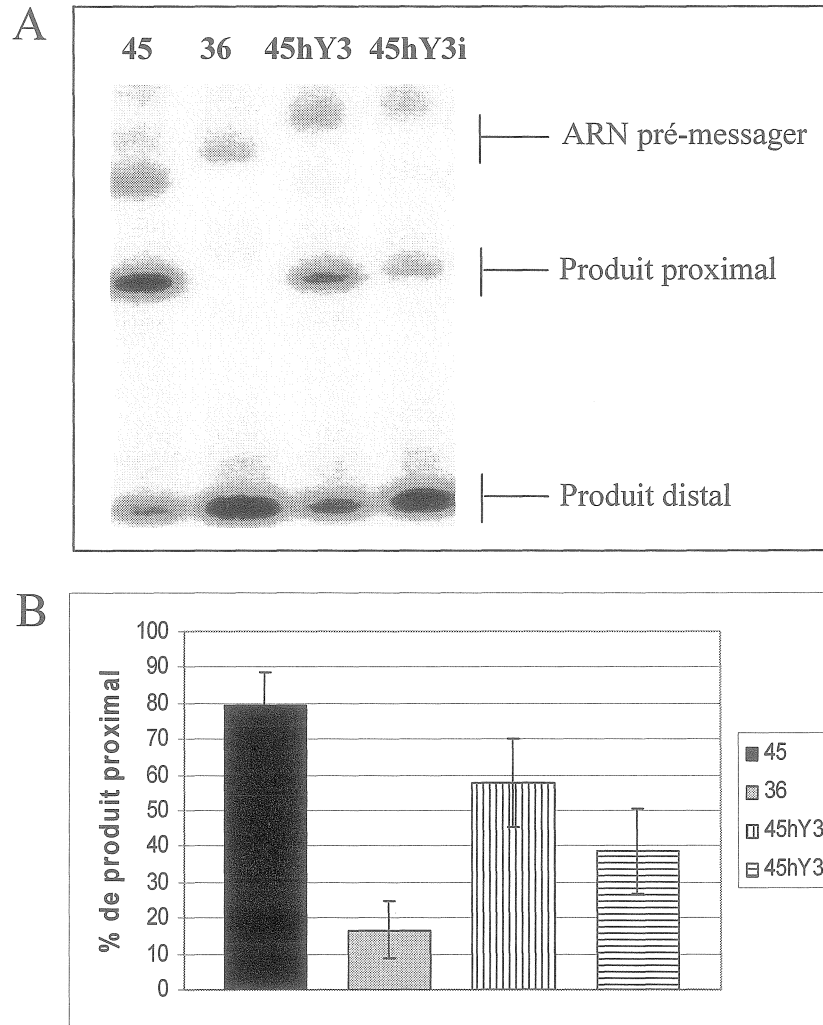


Figure 21 : Modulation de l'épissage *in vitro* du mini-gène hnRNP A1 contenant une séquence hY3 en aval de l'exon alternatif.

(A) Réactions d'épissage *in vitro* analysées par RT-PCR. Les produits issus du PCR ont ensuite été analysés sur un gel de polyacrylamide 5%. Les indications en haut de chaque piste indiquent les constructions utilisées (voir Figure 8).

(B) Comparaison des pourcentages de produit d'épissage proximal. Les résultats correspondent à la moyenne de 5 expériences indépendantes.

l'exon alternatif 7B. Dans le cas du plasmide 45hY3i, la séquence hY3 a été introduite dans l'orientation opposée à celle du mini-gène hnRNP A1. Lorsqu'on positionne une séquence hY3, dans l'orientation opposée, en aval de l'exon 7B, on observe un déplacement important du choix du site 5' proximal vers le site 5' distal (Figure 21). L'épissage de cet ARN pré-messager se fait maintenant à 39% au site 5' proximal, ce qui est un changement significatif comparativement au contrôle 45 (79%). La présence d'une séquence hY3 positionnée en aval de l'exon alternatif et dans la même orientation que le mini-gène modifie également la sélection du site d'épissage. Dans ce cas, le produit proximal est obtenu dans une proportion de 58%. Lorsque comparée avec le contrôle 45, la différence est limitée et peut-être pas significative. Donc, dans le cas où une séquence hY3 est présente en aval de l'exon alternatif, on observe une modulation de l'épissage favorisant le site d'épissage 5' distal; modulation qui est plus importante lorsque la séquence n'est pas dans la même orientation que le mini-gène.

4.12.1.2 Séquence hY3 en amont de l'exon alternatif

Le plasmide 45hY3.1 a été construit en insérant la séquence d'ADN correspondant à l'ARN hY3 à 30 nucléotides en aval de l'exon 7, au niveau du site de restriction *BamHI*. Lorsqu'on positionne une séquence hY3 en amont de l'exon 7B et dans la même orientation que le mini-gène (45hY3.1), la sélection du site d'épissage est peu modifiée (Figure 22). Dans ce cas, le produit proximal est obtenu dans une proportion de 70%, comparativement à 79% pour le contrôle 45, ce qui ne constitue

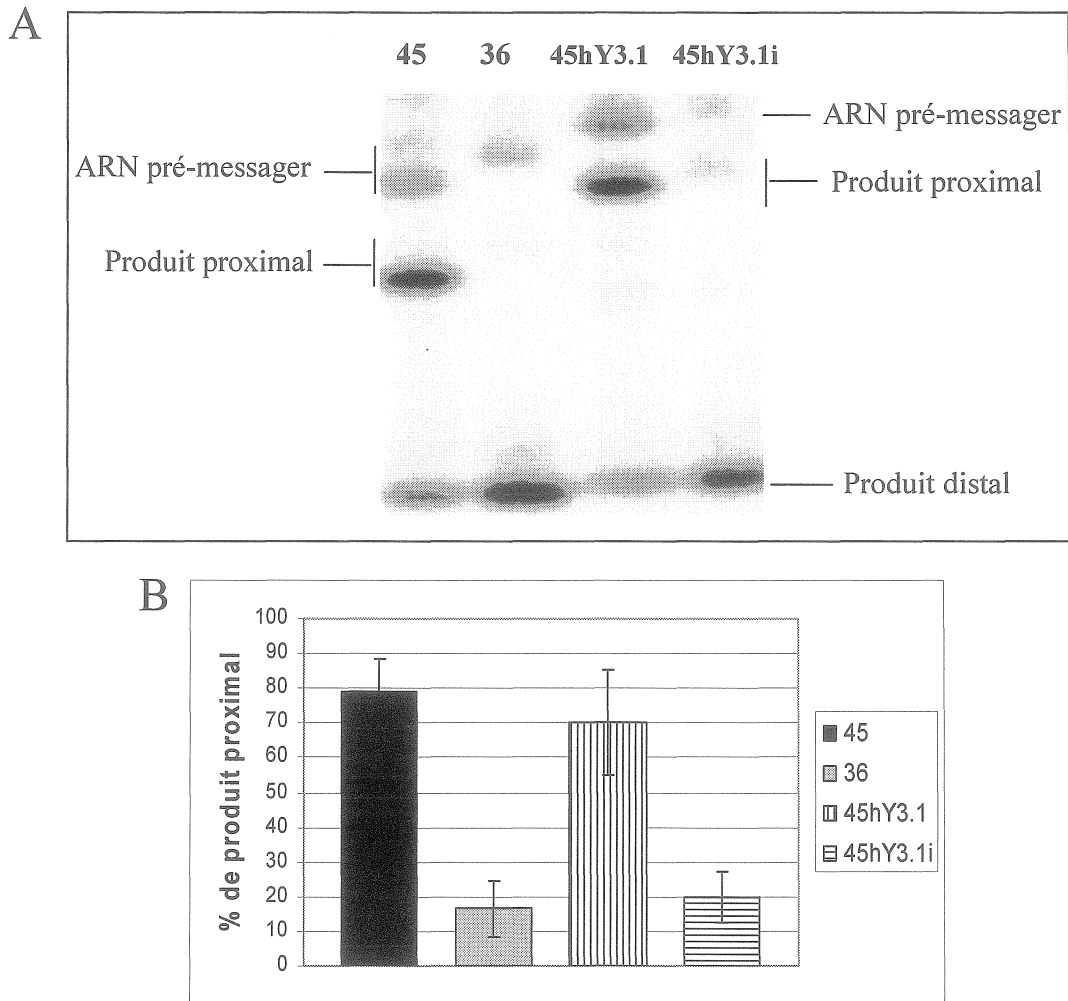


Figure 22 : Modulation de l'épissage *in vitro* du mini-gène hnRNP A1 contenant une séquence hY3 en amont de l'exon alternatif.

(A) Réactions d'épissage *in vitro* analysées par RT-PCR. Les produits issus du PCR ont ensuite été analysés sur un gel de polyacrylamide 5%. Les indications en haut de chaque piste indiquent les constructions utilisées (voir Figure 8).

(B) Comparaison des pourcentages de produit d'épissage proximal. Les résultats correspondent à la moyenne de 5 expériences indépendantes.

pas une différence significative. Pour la construction du plasmide 45hY3.1i, la séquence hY3 a été introduite dans l'orientation opposée à celle du mini-gène hnRNP A1. Dans ce cas, la présence de la séquence hY3 déplace de façon importante le choix du site 5' proximal vers le site 5' distal (Figure 22). L'épissage de cet ARN pré-messager se fait maintenant à 20% au site 5' proximal, ce qui constitue un changement significatif. Donc, comme observé dans le cas où une séquence hY3 est présente en aval de l'exon alternatif, la présence de la séquence hY3 en amont de l'exon alternatif déplace l'épissage de façon plus marquée vers le site d'épissage 5' distal lorsque la séquence n'est pas dans la même orientation que le mini-gène.

4.12.2 Les effets sur l'épissage de la séquence hY3 intronique sont constants d'un extrait nucléaire à un autre.

La modulation de l'épissage alternatif par des séquences hY3 est observée dans un système *in vitro* utilisant un extrait nucléaire de cellules HeLa. Afin de s'assurer de la reproductibilité des effets observés, nous avons effectué des expériences d'épissage *in vitro* en utilisant 3 extraits nucléaires de cellules HeLa préparés indépendamment. Ces extraits sont issus de la même lignée de cellules, mais n'ont pas été préparés au même moment. Dans ces extraits, des variations au niveau de la concentration de certains facteurs impliqués dans l'épissage peuvent entraîner des différences dans les patrons d'épissage observés *in vitro*. Ce phénomène a déjà été observé dans le laboratoire du Dr Chabot.

Dans les 3 extraits nucléaires sélectionnés, l'ARN pré-messager de base, transcrit à partir du plasmide 45, est épissé principalement au site d'épissage 5' de l'exon alternatif 7B. Le site 5' proximal est ainsi utilisé dans 91, 76 et 87% des cas (Figure 23). Les sites de liaisons de haute affinité pour la protéine hnRNP A1 de part et d'autre de l'exon 7B (plasmide 36) modifient toujours la sélection du site d'épissage 5' pour favoriser le site 5' distal. Dans ce cas, le site d'épissage 5' proximal n'est plus utilisé que dans 13, 16 et 30% des cas. Comme observé précédemment, la présence d'une séquence hY3 en aval de l'exon 7B (45hY3) favorise le site 5' proximal dans 75, 43 et 67% des cas, tandis que dans la situation où la séquence hY3 est dans l'orientation opposée (45hY3i), le produit proximal est obtenu dans une proportion de 56, 20 et 42%. Lorsqu'on compare l'épissage d'un même ARN pré-messager dans les 3 extraits nucléaires, on observe des différences notables. Par exemple, le plasmide 45hY3i donne 56% de produit proximal dans l'extrait NE278, comparativement à 20% dans l'extrait NE272. Par contre, peu importe l'extrait nucléaire, une séquence hY3 présente en aval de l'exon alternatif 7B module toujours l'épissage en favorisant le site d'épissage 5' distal et cette modulation est plus importante lorsque la séquence n'est pas dans la même orientation que le mini-gène. Les mêmes résultats sont obtenus lorsque la séquence hY3 est positionnée en amont de l'exon 7B. Donc, les effets sur l'épissage de la séquence hY3 intronique sont constants d'un extrait nucléaire à un autre.

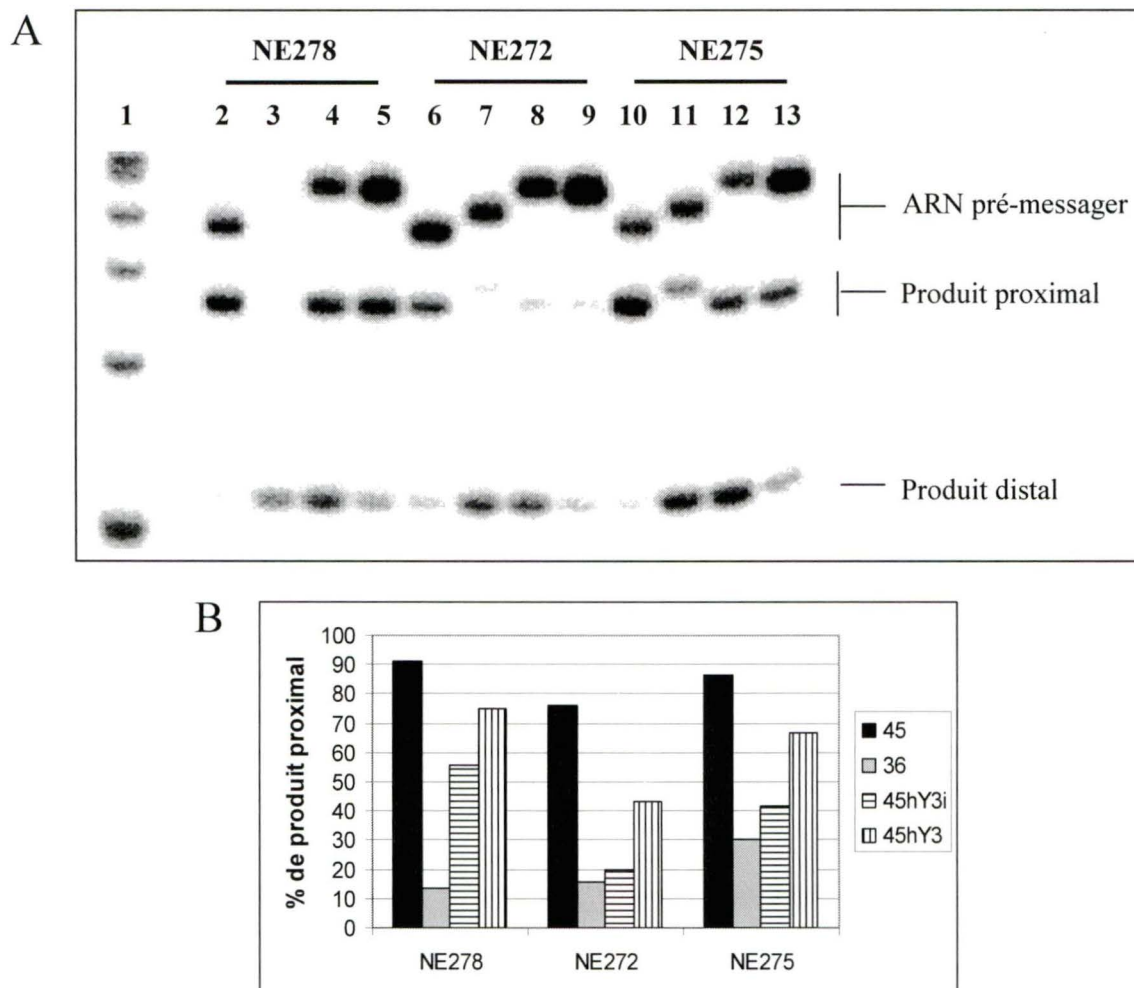


Figure 23 : Effets constants de la séquence hY3 intronique sur l'épissage avec différents extraits nucléaires.

(A) Réactions d'épissage *in vitro* analysées par RT-PCR. Les produits issus du PCR ont ensuite été analysés sur un gel de polyacrylamide 5%. La piste 1 contient un marqueur de poids moléculaire de 100 pb. Cette expérience est réalisée avec 3 extraits nucléaires de cellules HeLa préparés indépendamment (NE 278, NE 272 et NE 275). Le plasmide 45 a été utilisé pour les réactions des pistes 2, 6, 10, le plasmide 36 pour les pistes 3, 7, 11, le plasmide 45hY3i pour les pistes 4, 8, 12 et le plasmide 45hY3 pour les pistes 5, 9, 13.

(B) Comparaison des pourcentages de produit d'épissage proximal.

4.12.3 La modulation de l'épissage ne dépend pas d'un facteur *trans* : épissage avec ARN hY3 compétiteur.

L'épissage alternatif, ainsi que l'épissage constitutif, sont des processus cellulaires fort complexes qui impliquent de nombreux facteurs. Les différentes étapes de l'épissage font interagir ensemble plusieurs protéines et ARNs. Les principaux points de contrôle se situent au niveau de la reconnaissance des sites d'épissage.

Dans le but d'éclaircir le mécanisme par lequel une séquence hY3 dans un intron de l'unité d'épissage alternatif hnRNP A1 peut moduler l'épissage, des réactions d'épissage *in vitro* ont été réalisées en présence d'ARN hY3 compétiteur. De cette façon, si la modulation observée dépend de l'action de facteurs *trans*, c'est-à-dire de protéines ou d'ARNs interagissant directement avec la séquence hY3 intronique, la préincubation de l'extrait nucléaire avec cette même séquence devrait diminuer la disponibilité des facteurs et ainsi atténuer l'effet sur l'épissage.

Utilisant environ 2 fmol d'ARN pré-messager par réaction d'épissage, l'ajout de 75, 200 et 1000 fmol de compétiteur correspond à 37.5x, 100x et 500x la quantité d'ARN à épisser. Comme contrôle du système, avant l'ajout de l'ARN pré-messager transcrit à partir du plasmide 36, une quantité croissante d'oligonucléotide composé de plusieurs sites de liaisons pour hnRNP A1 (TS10) est préincubé avec le mélange d'épissage. Cet oligonucléotide contient 10 copies de la séquence GGGTTA. Dans ce cas, l'oligonucléotide, en liant la protéine hnRNP A1, va amoindrir l'effet des séquences de liaison à A1 introniques et déplacer la sélection du site d'épissage vers

le site 5' proximal. En effet, on observe la proportion d'utilisation du site d'épissage 5' proximal passer de 27% à 81% (Figure 24) ou de 6% à 74% (Figure 25). Lorsqu'une quantité croissante d'ARN hY3 (Figure 24) ou d'ARN hY3 antisens (Figure 25) est ajoutée aux réactions d'épissage de l'ARN pré-messager de base (45), on remarque une légère diminution de l'utilisation du site d'épissage 5' proximal. Cette diminution ne dépasse toutefois pas 15% et semble déjà être maximale même à la plus faible concentration d'ARN compétiteur. Une modulation de cette ampleur ne sera donc pas considérée importante, puisqu'elle est obtenue avec le contrôle négatif 45.

Des quantités croissantes d'ARN hY3 ont été ajoutées à la réaction d'épissage de l'ARN pré-messager 45hY3 afin de vérifier si l'ARN hY3 peut compétitionner pour des facteurs actifs. L'ajout d'ARN hY3 modifie légèrement le choix du site d'épissage 5', l'effet maximum étant encore une fois observé à faible concentration d'ARN compétiteur. La proportion du produit proximal passe de 51% à 34%, ce qui constitue un changement de 17%. Par contre, cette différence est faible lorsque comparée à celle obtenue avec le contrôle de base 45 (15%). De la même façon, l'ajout d'ARN hY3 antisens compétiteur à la réaction d'épissage de l'ARN pré-messager 45hY3i ne modifie que très légèrement le choix du site 5' (32% à 27%). Donc, la modulation de l'épissage observée dans le cas où une séquence hY3 est présente en aval de l'exon alternatif ne semble pas dépendre de l'action d'un facteur en *trans*.

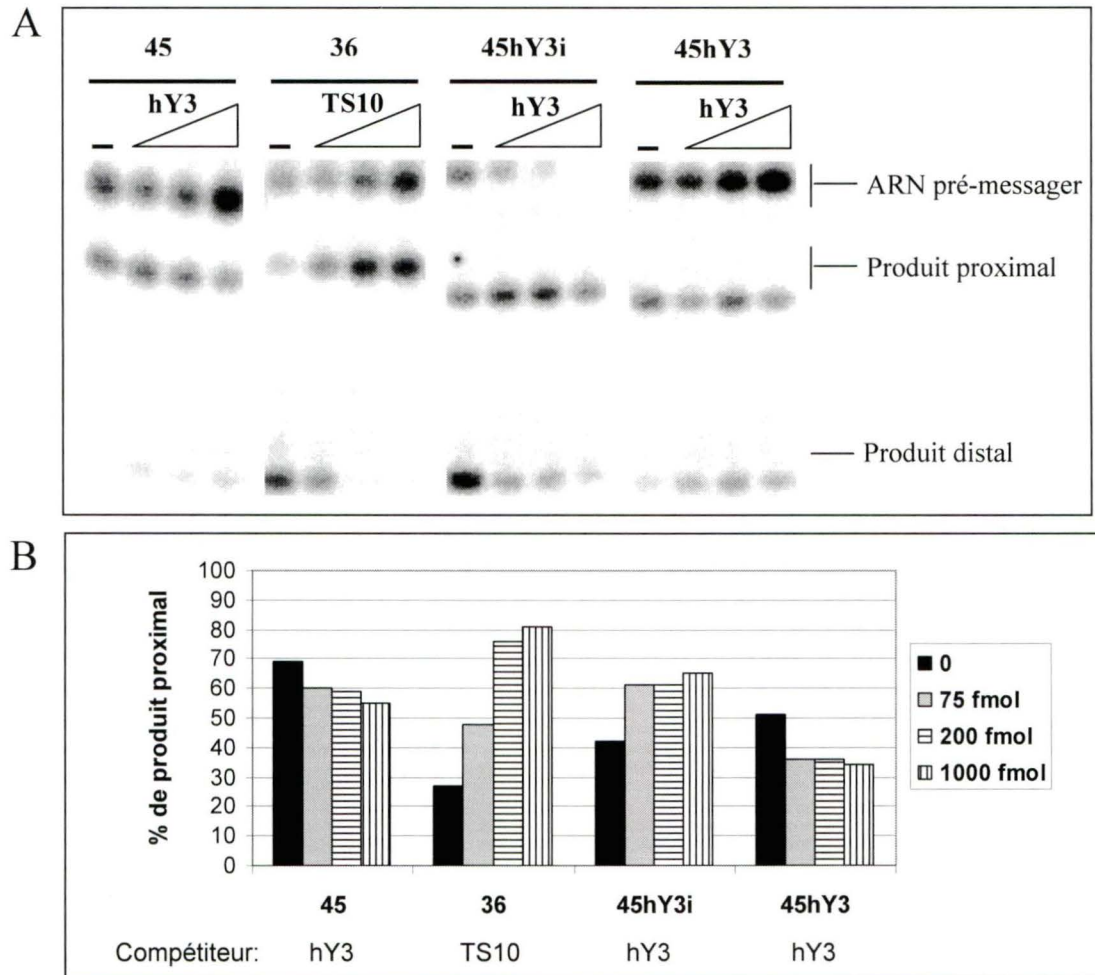


Figure 24 : Modulation de l'épissage *in vitro* du mini-gène hnRNP A1 avec ARN hY3 compétiteur.

(A) Réactions d'épissage *in vitro* analysées par RT-PCR. Les produits issus du PCR sont analysés sur un gel de polyacrylamide 5%. Les indications en haut de chaque panneau indiquent les constructions utilisées (voir Figure 8) ainsi que la nature et les variations de quantité des compétiteurs utilisés. Les réactions d'épissage sont effectuées dans un extrait nucléaire de cellules HeLa préincubé pendant 10 min avec des quantités croissantes (0, 75, 200, 1000 fmol) de compétiteur, soit d'ARN hY3, ou d'oligonucléotide contenant des sites de liaison pour hnRNP A1 (TS10).

(B) Comparaison des pourcentages de produit d'épissage proximal.

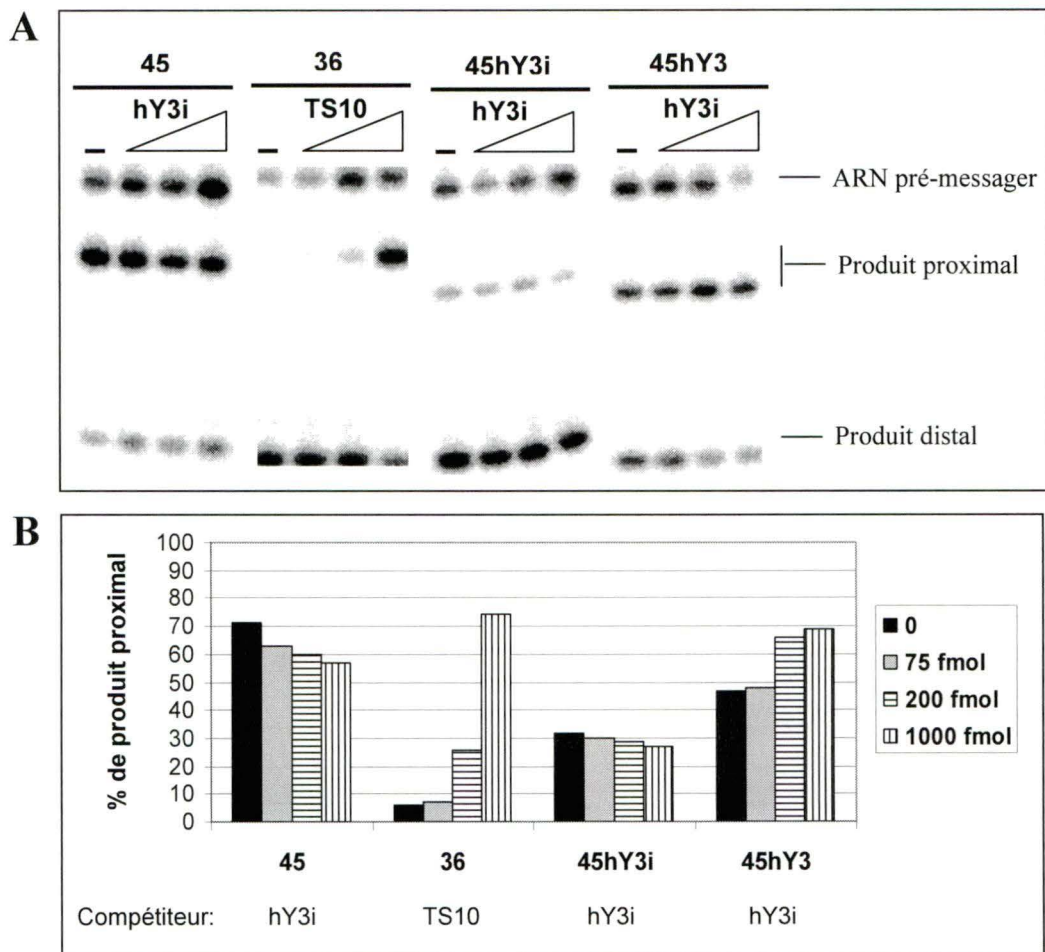


Figure 25 : Modulation de l'épissage *in vitro* du mini-gène hnRNP A1 avec ARN hY3 antisens compétiteur.

(A) Réactions d'épissage *in vitro* analysées par RT-PCR. Les produits issus du PCR sont analysés sur un gel de polyacrylamide 5%. Les indications en haut de chaque piste indiquent les constructions utilisées (voir Figure 8) ainsi que la nature et les variations de quantité des compétiteurs utilisés. Les réactions d'épissage sont effectuées dans un extrait nucléaire de cellules HeLa préincubé pendant 10 min avec des quantités croissantes (0, 75, 200, 1000 fmol) de compétiteur, soit d'ARN hY3 antisens (hY3i), ou d'oligonucléotide contenant des sites de liaison pour hnRNP A1 (TS10). (B) Comparaison des pourcentages de produit d'épissage proximal.

D'autre part, lorsqu'on ajoute une quantité croissante d'ARN hY3 à la réaction d'épissage de l'ARN pré-messager contenant la séquence dans l'orientation opposée (45hY3i), on observe un déplacement important de l'épissage vers le site 5' proximal. L'utilisation du site 5' proximal passe de 42% à 65%, ce qui est un effet majeur lorsque comparée au contrôle de base (45) qui passe de 69% à 55%. De même, l'ajout d'une quantité croissante d'ARN hY3 antisens (hY3i) à la réaction d'épissage de l'ARN pré-messager contenant la séquence opposée (45hY3) favorise aussi grandement le site d'épissage 5' proximal. La proportion de produit proximal passe ainsi de 47% à 69% (Figure 25). L'ampleur de ces effets est comparable à celle observée lors de l'ajout de compétiteur TS10 à la réaction d'épissage de l'ARN pré-messager 36 (28% proximal à 48% proximal). Donc, la modulation de l'épissage par une séquence hY3 positionnée en aval de l'exon 7B est renversée par la présence dans le mélange d'épissage d'une séquence complémentaire.

Les mêmes expériences de compétition ont été réalisées en utilisant les plasmides 45hY3.1 et 45hY3.1i, dans lesquels une séquence hY3 est positionnée en amont de l'exon 7B (Figure 26). Dans ces expériences, l'ajout d'ARN hY3 à la réaction d'épissage de l'ARN pré-messager transcrit à partir du plasmide contrôle 45 déplace encore légèrement l'épissage vers le site d'épissage 5' distal, la sélection du site proximal passant de 75% à 61%. Lorsque l'oligonucléotide TS10 est présent dans la réaction d'épissage du plasmide 36, on assiste toujours à un retour de l'épissage vers le site 5' proximal, dont l'utilisation passe de 11% à 36%. L'ajout d'ARN hY3 à la réaction d'épissage de l'ARN pré-messager 45hY3.1 ne modifie pratiquement pas la sélection du site d'épissage 5' (68% proximal à 71% proximal). Par contre,

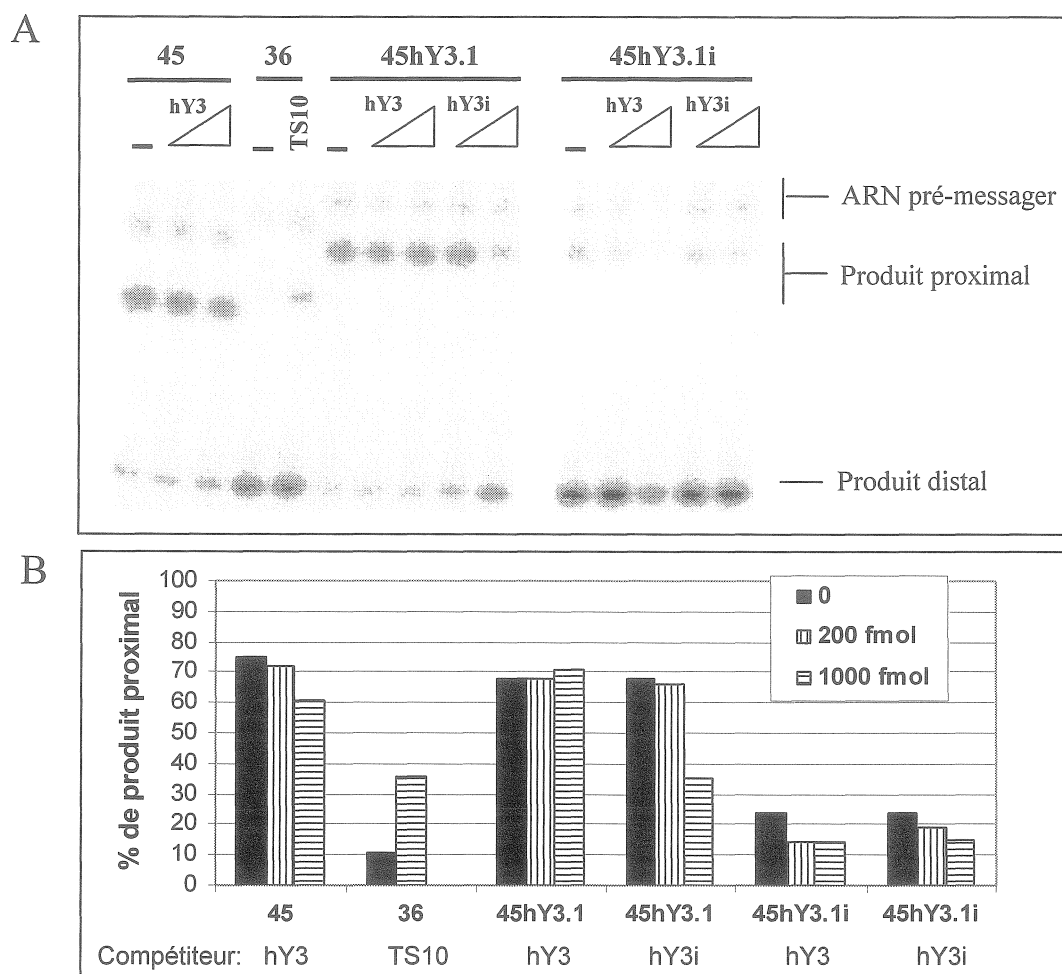


Figure 26: Modulation de l'épissage *in vitro* du mini-gène hnRNP A1 avec ARN hY3 ou ARN hY3 antisens compétiteur.

(A) Réactions d'épissage *in vitro* analysées par RT-PCR. Les produits issus du PCR sont analysés sur un gel de polyacrylamide 5%. Les indications en haut de chaque panneau indiquent les constructions utilisées (voir Figure 8) ainsi que la nature et les variations de quantité des compétiteurs utilisés. Les réactions d'épissage sont effectuées dans un extrait nucléaire de cellules HeLa préincubé pendant 10 min avec des quantités croissantes (0, 200, 1000 fmol) de compétiteur, soit d'ARN hY3, d'ARN hY3 antisens (hY3i) ou d'oligonucléotide contenant des sites de liaison pour hnRNP A1 (TS10). (B) Comparaison des pourcentages de produit d'épissage proximal. Les indications en bas du graphique indiquent les constructions (voir Figure 8) et les compétiteurs utilisées.

considérant que la présence d'ARN hY3 dans la réaction d'épissage du contrôle 45 déplace l'épissage vers le site 5' distal de 14%, on peut calculer un déplacement net de 17% en direction du site 5' proximal. Donc, dans cette situation, il semblerait y avoir une légère compétition pour un facteur actif en *trans*. D'autre part, l'utilisation du site d'épissage 5' proximal passe de 24% à 15% lorsqu'une quantité croissante d'ARN hY3 antisens est ajoutée à la réaction d'épissage de l'ARN pré-messager 45hY3.1i. Ce changement est peu important lorsque comparé au changement observé précédemment pour le contrôle de base (45) qui passait de 71% à 57% (Figure 26). Donc, dans ce cas, l'ARN hY3 antisens ajouté ne compétitionne avec aucun facteur.

L'ajout d'ARN hY3 à la réaction d'épissage de l'ARN pré-messager contenant une séquence hY3 dans l'orientation opposée déplace légèrement l'épissage vers le site d'épissage 5' distal, la proportion d'utilisation du site proximal passant de 24% à 14%. Cette différence de 10% est faible lorsque comparée à celle obtenue avec le contrôle de base 45 (14%). Finalement, l'épissage de la construction 45hY3.1 est influencé par l'ajout d'ARN hY3 antisens, de sorte que le pourcentage de produit proximal passe de 68% à 35%. Cette variation de 33% est notable lorsque comparée au contrôle 45 (14%). Donc, dans ces 2 dernières situations, la modulation de l'épissage par une séquence hY3 positionnée en amont de l'exon 7B n'est pas amoindrie par la présence dans la réaction d'épissage d'une séquence complémentaire. La modulation est même fortement accentuée dans un des cas.

4.13 Modèle d'épissage alternatif Bcl-x

Les séquences hY3 introniques influencent la sélection du site d'épissage 5', importante pour l'inclusion ou l'exclusion de l'exon alternatif 7B dans le modèle hnRNP A1. Par contre, cette modulation ne semble pas dépendre de l'action d'un facteur en *trans*. Afin de vérifier si la modulation par les séquences hY3 est spécifique au modèle hnRNP A1, ou si elle peut être observée dans d'autres modèles, des séquences hY3 ont été insérées dans un autre modèle d'épissage. Cette nouvelle unité d'épissage alternatif est dérivée du gène Bcl-x et est aussi utilisée dans le laboratoire du Dr Chabot.

Ce modèle a été développé pour étudier l'épissage de l'ARN pré-messager du gène Bcl-x, plus spécifiquement au niveau de la régulation de l'épissage de l'exon 2. Il est constitué d'une partie des séquences des exons 2 et 3. L'exon 2 contient 2 sites d'épissage 5' en compétition pour un site d'épissage 3'. L'isoforme Bcl-xS est produit lorsque le site 5' le plus éloigné du site 3' est utilisé, tandis que l'isoforme Bcl-xL est produit lorsque le site 5' le plus près du site 3' est utilisé (Figure 9). L'intron séparant les 2 exons, initialement d'une taille de plus de 55 kb, a été amputé jusqu'à une taille d'environ 1,4 kb. Pour vérifier l'influence sur l'épissage de séquences hY, il est possible de positionner ces séquences dans l'intron et d'observer la modulation du choix du site d'épissage 5'.

4.13.1 Les séquences hY3 introniques influencent peu l'épissage *in vitro* du mini-gène Bcl-x.

L'ARN pré-messager transcrit à partir de l'unité d'épissage alternatif de base (plasmide Bclx), est épissé majoritairement au site d'épissage 5' produisant l'isoforme Bcl-xL. Ce site d'épissage 5' est ainsi sélectionné dans 92% des cas, tandis que le site d'épissage produisant l'isoforme Bcl-xS n'est utilisé que dans 8% des cas (Figure 27). Le plasmide Bclx3.1 a été construit en remplaçant le segment *BsmBI* - *BstEII* par la séquence d'ADN correspondant à l'ARN hY3 et le plasmide Bclx3.2 en remplaçant le segment *BsaBI* - *BseRI*. L'épissage *in vitro* des ARNs pré-messagers issus de ces 2 plasmides produit l'isoforme Bcl-xL dans des proportions de 91% et 89%, ce qui constitue une très faible différence par rapport au contrôle de base. L'effet des séquences hY3 orientées dans le sens contraire au mini-gène (Bclx3.1i et Bclx3.2i) est légèrement plus important, avec des proportions de l'isoforme Bcl-xL de 85% et 87%. Ces différences sont significatives, mais leur ampleur reste faible. Donc, les séquences hY3 introniques influencent peu l'épissage *in vitro* du mini-gène Bcl-x.

4.13.2 Les séquences hY3 introniques influencent peu l'épissage *in vivo* du mini-gène Bcl-x

L'épissage *in vitro* de l'unité d'épissage alternatif Bcl-x semble peu affecté par la présence de séquences hY3 introniques, peu importe la distance des sites d'épissage. Par contre, *in vivo*, l'épissage de cette unité diffère considérablement.

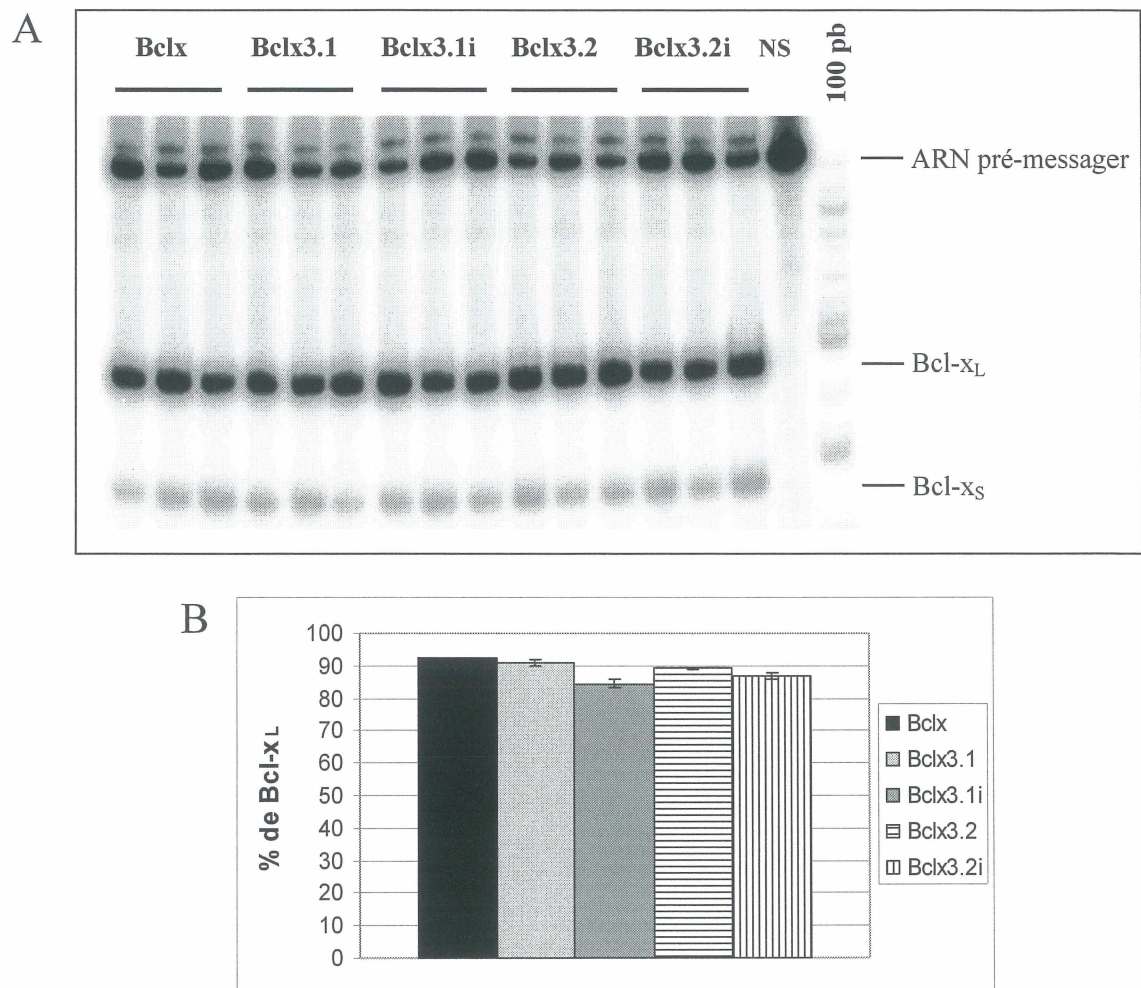


Figure 27 : Épissage *in vitro* du mini-gène Bcl-x contenant des séquences hY3 introniques.

(A) Réactions d'épissage *in vitro* analysées par RT-PCR. Les produits issus du PCR sont analysés sur un gel de polyacrylamide 4%. Les réactions d'épissage sont effectuées en triplicata pour chaque construction testée. Les indications en haut de chaque piste indiquent les constructions utilisées (voir Figure 10). La piste identifiée NS représente une réaction de RT-PCR effectuée sur un ARN pré-messager transcrit, mais non épissé.

(B) Comparaison des pourcentages de l'isoforme Bcl-xL. Les résultats correspondent à la moyenne de 3 réactions.

Effectivement, *in vivo*, cette unité d'épissage alternatif est épissée majoritairement au site 5' produisant l'isoforme Bcl-xS (Figure 28). *In vivo*, la proportion de Bcl-xL diminue à 19% dans les cellules HeLa et à 27% dans les cellules ECR 293. Lorsqu'une séquence hY3 est insérée dans l'intron, on observe de légers changements dans la sélection du site d'épissage 5'. Par contre, les effets observés sont de faible amplitude, comparativement au contrôle de base, peu importe l'orientation des séquences ou la distance du site d'épissage 5'. Les différences n'atteignent pas 10%. Donc, les séquences hY3 introniques influencent peu l'épissage *in vivo* du mini-gène Bcl-x.

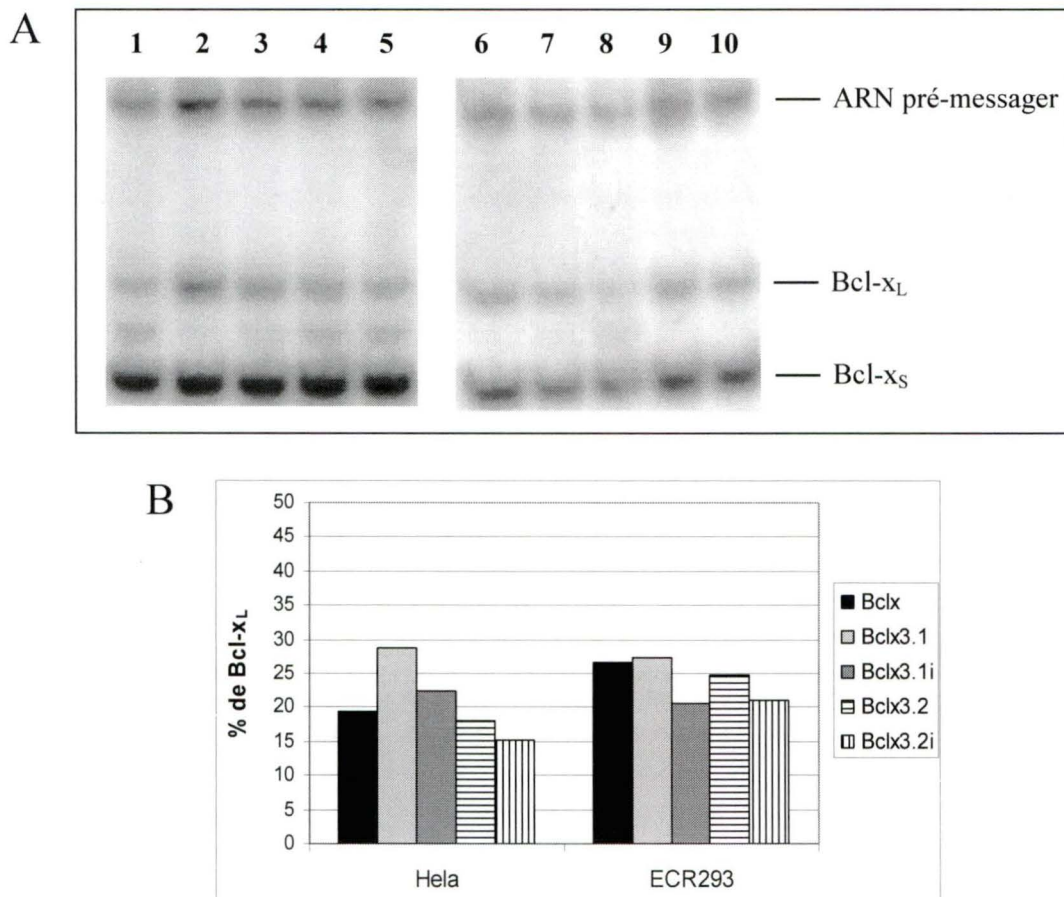


Figure 28 : Épissage *in vivo* du mini-gène Bcl-x contenant des séquences hY3 introniques.

(A) Réactions d'épissage *in vivo* analysées par RT-PCR. Les produits issus du PCR ont ensuite été analysés sur un gel de polyacrylamide 4%. Le plasmide Bclx a été utilisé pour les réactions d'épissage des pistes 1 et 6, Bclx3.1 pour les pistes 2 et 7, Bclx3.1i pour les pistes 3 et 8, Bclx3.2 pour les pistes 4 et 9, et Bclx3.2i pour les pistes 5 et 10.

Panneau de gauche : cellules HeLa

Panneau de droite : cellules ECR 293

(B) Comparaison des pourcentages de l'isoforme Bcl-xL produit.

5. DISCUSSION

5.1 Levure et RoBP1

L'identification des fonctions cellulaires des RNPs Ro pourrait aider à la compréhension de la réponse autoimmune dirigée contre ces particules et leurs composantes dans certaines maladies du tissu conjonctif. Dans cette optique, les travaux du Dr Pascal Bouffard ont permis l'identification et la caractérisation de RoBP1, un nouveau partenaire de la RNP Ro. Afin de mieux caractériser cette protéine et d'en définir la fonction, Jim Boulanger a tenté d'identifier l'homologue fonctionnel de RoBP1 chez *Saccharomyces cerevisiae* en créant une souche de levure dépendante de l'expression de RoBP1 humaine. Par test de complémentation dans cette souche de levure, deux fragments des chromosomes IV et XI avaient été isolés d'une librairie génomique de *S. cerevisiae* et semblaient pouvoir remplacer RoBP1 dans la levure dépendante de RoBP1. Sur le chromosome XI, le gène CAF4 avait été identifié (Mémoire de maîtrise, Jim Boulanger, 2000, Faculté de Médecine, Université de Sherbrooke).

Avant de poursuivre avec la définition plus précise des homologues fonctionnels de RoBP1, la première étape du projet était donc la confirmation des données obtenues par Jim Boulanger. La souche de levure dépendante de RoBP1 utilisée est issue d'un rétrocroisement avec la souche haploïde BY4705 MATa. Les colonies de couleur blanche sur gélose YEPD et la croissance sur gélose Yc-Ura indiquent la conservation du plasmide pCu4X6/RoBP1 contenant les marqueurs

ADE2 et URA3. De plus, l'importance de ce plasmide est confirmée par une croissance très rare de cette souche sur gélose 5'FOA, donnant seulement quelques micro-colonies. Mais même la souche contrôle BY4704 peut donner à l'occasion quelques micro-colonies après 5-7 jours de croissance sur ce milieu. Donc, tout porte à croire que la souche dépendante de RoBP1 est bel et bien dépendante de RoBP1.

Par contre, cette souche produit des colonies de diamètre inférieur à celles produites par la souche sauvage BY4705 sur milieu solide. Ce phénotype peut se retrouver chez certaines levures qui ne peuvent utiliser les sucres que par fermentation, à cause de défauts de la fonction mitochondriale. C'est le cas pour la levure dépendante de RoBP1 qui est incapable de croître sur un milieu où la seule source de carbone est non-fermentable (YEPG). Ce type de défaut peut provenir de larges délétions dans le génome mitochondrial ou même d'une absence totale de génome mitochondrial. Dans ces 2 cas, les levures sont désignées ρ^- et ρ^0 , respectivement (GOLDRING et al., 1970). Des mutations altérant la fonction de certains gènes encodés par le génome nucléaire peuvent également produire des levures petites, nommées *pet⁻*. De plus, il est à noter que les levures petites, incapables de métabolisme aérobie, ne peuvent oxyder le composé responsable de la formation du pigment rouge et montrent donc une couleur blanche même en absence du marqueur ADE2. Ces mutations proviennent vraisemblablement de la mutagenèse à l'EMS effectuée lors de la création de la souche dépendante de RoBP1.

La souche de levure dépendante de RoBP1 initiale créée par Jim Boulanger avait déjà cette déficience respiratoire comme le montre son incapacité à croître sur le

milieu YEPG. Même les autres spores issues du premier rétrocroisement avec la souche BY4705 sauvage ayant conservé la dépendance à RoBP1 sont toujours petites.

Dans le but de récupérer une souche dépendante de RoBP1 dont les mitochondries sont fonctionnelles, un deuxième rétrocroisement de la levure dépendante de RoBP1 avec la levure BY4705 sauvage a été effectué. Ce croisement avec la souche parentale peut permettre de se débarrasser des mutations indésirables amenées lors de la mutagenèse. Toutefois, aucune des spores obtenues après microdissection des tétrades n'a à la fois conservé la dépendance à RoBP1 et récupéré des fonctions mitochondriales normales. Toutes les spores dépendantes de RoBP1 sont également petites. Il semble donc y avoir un lien entre ces deux phénotypes. Ce lien peut être physique ou fonctionnel. Les mutations produisant ces deux phénotypes peuvent être présentes dans des gènes situés à une courte distance sur le même chromosome. Cette proximité rend plus difficile la séparation de ces deux gènes par recombinaison lors de la méiose. D'un autre côté, le ou les homologues fonctionnels de RoBP1 chez la levure pourraient aussi avoir simplement un rôle au niveau du métabolisme aérobie.

La déficience respiratoire de la souche dépendante de RoBP1, entraînant la production de petites colonies, rend l'interprétation de la croissance sur le milieu 5'FOA difficile. En effet, la très faible croissance sur ce milieu pourrait être expliquée par la déficience respiratoire plutôt que par l'impossibilité pour cette souche de se débarrasser du plasmide contenant RoBP1. Une combinaison de ces deux explications est également possible. Une levure montrant une dépendance seulement

partielle à RoBP1 combinée avec des fonctions mitochondriales altérées pourrait produire le même phénotype de croissance très faible sur milieu 5'FOA.

Des tests de complémentation ont aussi été effectués afin de préciser l'identification des homologues fonctionnels de RoBP1 amorcée par Jim Boulanger. Les résultats sont ambigus, certains plasmides permettant quelquefois la production de micro-colonies. Encore une fois, l'interprétation difficile de la croissance sur le milieu 5'FOA empêche une complémentation claire. Même les levures contenant le plasmide pGADGH/RoBP1, considéré comme un contrôle positif, ne montrent pas de croissance convaincante sur ce milieu.

Les tests de complémentation préliminaires effectués par Jim Boulanger avec la levure dépendante de RoBP1 initiale suggéraient que des mutations dans 2 gènes pouvaient être responsables de la dépendance à RoBP1. Ces travaux avaient identifié CAF4 sur le chromosome XI. En analysant les fonctions confirmées ou hypothétiques des gènes du chromosome IV, le gène UBC5 semblait être le meilleur candidat pour le deuxième gène. Donc, si ces deux gènes sont responsables de la dépendance à RoBP1, leur délétion individuelle ou combinée devrait être létale pour la levure.

Lors de la méiose d'une souche diploïde hétérozygote pour la délétion d'un gène essentiel il y a production de 2 spores viables et 2 spores non-viables. Cette ségrégation est obtenue avec la souche diploïde hétérozygote pour la délétion du gène UBC5, ce qui suggère que ce gène est essentiel. Par contre, plusieurs spores peuvent croître sur le milieu restrictif Yc-Leu, montrant la présence du marqueur LEU2 et l'absence du gène UBC5. Donc UBC5 n'est pas un gène essentiel. La même analyse est effectuée avec les résultats obtenus avec le gène CAF4. La viabilité des souches

contenant les délétions des gènes CAF4 ou UBC5 est aussi vérifiée sur le site internet SGD (*Saccharomyces* Genome Database; www.yeastgenome.org). Par contre, la ségrégation 2 : 2 (vie:mort) montre quand même la présence d'une mutation létale dans ces 2 souches, ce qui est difficilement explicable. Cette mutation pourrait être survenue lors de la transformation des fragments de délétion, qui est un procédé reconnu comme étant mutagénique.

Lors de la méiose, la ségrégation des allèles de 2 gènes synthétiquement létaux produit 3 spores viables et 1 spore non-viable. Les tétrades obtenues après sporulation de la souche diploïde hétérozygote pour la délétion des gènes UBC5 et CAF4 montrent une ségrégation 2 : 2. De plus, plusieurs spores peuvent croître sur le milieu restrictif Yc-Leu-His, montrant la présence des marqueurs LEU2 et HIS3 et donc l'absence des gènes UBC5 et CAF4. Les gènes UBC5 et CAF4 ne sont donc pas synthétiquement létaux.

Un agent mutagène comme l'EMS va produire plusieurs mutations aléatoires dans tout le génome et cette accumulation de mutations peut affecter la croissance des levures. De plus, des mutations peuvent survenir dans des gènes directement impliqués dans la respiration. Donc, il est courant d'obtenir de petites colonies après une telle mutagénèse. Pour cette raison, il est de routine d'examiner la possibilité que la croissance ralentie provienne d'un défaut respiratoire en vérifiant la croissance sur gélose YEPG. Habituellement, lorsqu'on ne recherche pas spécifiquement un gène impliqué dans la respiration, ces clones de levure petites sont rejetés. Cette vérification n'avait pas été effectuée avec le seul clone positif « dépendant de

RoBP1 » obtenu. Finalement, plusieurs résultats montrés dans les travaux effectués préalablement ont été impossibles à reproduire et sont donc difficilement explicables.

5.2 Pseudogènes hY

L'existence de séquences homologues aux ARNs hY distribuées dans le génome humain est connue depuis de nombreuses années, mais leur localisation et leur caractérisation n'avaient pas été approfondies jusqu'à tout récemment. La recherche de séquences homologues aux gènes codant pour les ARNs hY dans le génome humain a identifié près de 1000 pseudogènes hY dispersés dans tous les chromosomes. Les critères de sélection étant relativement stricts, ce nombre est probablement une sous-estimation de la quantité réelle de pseudogènes dérivés des ARNs hY. Chacun des 4 ARNs hY a engendré des pseudogènes, bien que la plupart soient dérivés des ARNs hY1 et hY3. Environ 95% des pseudogènes hY trouvés chez l'humain sont également retrouvés chez le chimpanzé aux mêmes positions génomiques. Au contraire, le génome de la souris contient uniquement 27 pseudogènes Y et aucun n'est situé dans des régions conservées par rapport à l'humain. Le degré de conservation de la séquence des pseudogènes hY lorsque comparée aux séquences des gènes authentiques correspondants, suggèrent également que la plupart des pseudogènes Y sont apparus après la divergence des rongeurs et des primates. Cependant, cette explication n'exclut pas complètement la possibilité que la quantité inférieure de pseudogènes Y chez la souris résulte d'un taux de mutation plus élevé menant à une dégradation plus rapide des séquences rétrotransposées.

L'impact physiologique ou pathologique de la présence dans le génome humain d'une grande quantité de pseudogènes hY reste à être étudié. Près de 42% des pseudogènes hY ont été trouvés dans les régions introniques de gènes codant pour des protéines. Cette distribution, similaire à celle des éléments Alu, pourrait s'expliquer par une sélection positive en faveur d'une fonction. En effet, le rôle des éléments Alu introniques dans la génération de sites d'épissage alternatif et leur contribution dans la diversité des cellules et tissus est de plus en plus reconnu. Sachant que les ARNs hY sont associés à plusieurs protéines ayant des rôles dans l'épissage alternatif, il est possible que certaines de ces séquences introniques homologues aux ARNs hY puissent influencer l'épissage alternatif.

Une analyse attentive de la séquence des pseudogènes et des régions adjacentes révèle que la plupart des pseudogènes hY ne sont pas le résultat d'événements de duplication. Au contraire, plusieurs évidences suggèrent que les pseudogènes hY ont été générés par la machinerie de rétrotransposition de L1. Près de 70% de ces pseudogènes présentent des signatures claires de rétrotransposition par L1. Le système d'évaluation de ces signatures étant relativement strict, le nombre réel de pseudogènes hY issus d'un événement de rétrotransposition par L1 est probablement plus élevé. Certains éléments plus anciens, avec des TSDs dégénérés, peuvent avoir échappé à l'analyse. De plus, des événements récents observés expérimentalement peuvent présenter des TSDs aussi courts que 2 nt (DEWANNIEUX et al., 2003).

Les pseudogènes hY ont donc vraisemblablement été générés par la rétrotransposition directe des ARNs hY par la machinerie L1. Contrairement à ce qui

est observé pour certains ARNs U, il y a peu d'évidence montrant l'existence de pseudogènes hY chimériques. La plupart des éléments Alu retrouvés à proximité des pseudogènes hY semblent s'être insérés ultérieurement dans la queue poly(A) ou le TSD d'un pseudogène hY existant. Dans de rares cas, la séquence séparant le pseudogène hY de l'élément Alu est tellement courte que la possibilité de chimère ne peut être écartée.

L'identification dans le génome humain d'une grande quantité de pseudogènes résultant directement de la rétrotransposition des ARNs hY est intéressante. Le nombre de pseudogènes hY est significativement plus élevé que le nombre de pseudogènes dérivés de certains ARNs non-codants 100 à 1000 fois plus abondants dans la cellule, comme l'ARNr 5S et l'ARN U1. De plus, les ARNs hY n'ont pas de queue poly(A) à leur extrémité 3', ce qui est reconnu comme étant important pour une rétrotransposition efficace par L1 (DEWANNIEUX et al., 2003). Par contre, la rétrotransposition par L1 peut se produire même sans queue poly(A), mais à un niveau plus réduit (ROY-ENGEL et al., 2002).

Les RNPs Ro humaines ont des propriétés immunologiques et biochimiques distinctes qui sont déterminées par l'ARN hY qu'elles contiennent (BOIRE & CRAFT, 1990; GENDRON et al., 2001). L'association des ARN hY1 et hY3 avec des protéines spécifiques (PTB, hnRNP K) pourrait avoir contribué à l'abondance relative des pseudogènes dérivés de ces 2 ARNs. Au contraire, l'ARN hY5 semble n'avoir été la source que de quelques pseudogènes. Le fait que l'ARN hY5 soit une acquisition récente des primates et qu'il soit le seul associé spécifiquement à la

protéine RoBP1 peut avoir joué un rôle dans cette faible fréquence d'événements de rétrotransposition.

D'un autre côté, la fréquence élevée de mutations ponctuelles, introduites au niveau de l'ARN ou de l'ADN, à des positions ciblées dans les pseudogènes hY suggère que l'absence de liaison à des protéines comme Ro60 et La pourrait faire de ces ARNs hY mutés de meilleurs substrats pour la rétrotransposition. On ne peut cependant pas exclure la possibilité que la présence d'une trop grande quantité de pseudogènes hY fonctionnels soit néfaste pour la cellule. Ces séquences subiraient ainsi une pression de sélection négative et seraient éliminées du génome.

5.3 hY et épissage

L'étude de la distribution génomique des pseudogènes des ARNs hY a permis d'observer qu'une forte proportion se retrouve dans des séquences géniques, plus précisément, dans des introns. Nous avons localisé 23 pseudogènes hY situés à moins de 200 nucléotides d'un site d'épissage 5' ou 3'. Bien qu'il soit plutôt difficile d'évaluer avec précision la structure secondaire d'un ARN pré-messager, il est probable que les séquences hY3 présentes puissent adopter la structure en tige-boucle des ARN Y. Un repliement semblable a été observé pour quelques pseudogènes en utilisant le logiciel de modélisation *mfold* (MATHEWS et al., 1999; ZUCKER, 2003). Ces séquences introniques pourraient influencer l'épissage d'un exon adjacent, sachant qu'elles peuvent lier des protéines impliquées dans l'épissage. Afin de vérifier cette hypothèse, nous avons examiné l'effet de la présence de séquences

correspondant aux ARNs hY à proximité de sites d'épissage sur l'épissage alternatif de gènes modèles.

5.3.1 Modèle d'épissage alternatif hnRNP A1

Le premier modèle utilisé contient une unité d'épissage alternatif dérivée du gène hnRNP A1. L'épissage *in vitro* de cette unité se fait principalement au site 5' proximal. Sachant que dans le génome humain, l'orientation des pseudogènes hY par rapport aux gènes dans lesquels ils sont situés ne semble montrer aucune préférence, la séquence hY3 a été positionnée dans les 2 orientations par rapport au mini-gène hnRNP A1. Premièrement, la séquence d'ADN correspondant à l'ARN hY3 a été insérée à une distance de 30 nucléotides en aval de l'exon alternatif 7B. L'ajout d'une séquence hY3 à cette position modifie l'épissage *in vitro* de l'ARN pré-messager, ayant pour résultat une diminution de la proportion d'utilisation du site d'épissage 5' proximal. Cet effet semble plus marqué lorsque la séquence hY3 est dans l'orientation contraire au mini-gène. Deuxièmement, la séquence d'ADN correspondant à l'ARN hY3 a aussi été insérée en amont de l'exon alternatif, à une distance de 29 nucléotides en aval du site d'épissage 5' de l'exon 7. La présence d'une séquence hY3 à cette position modifie également l'épissage *in vitro* de l'ARN pré-messager de façon à générer une plus grande proportion de produit distal. Cette modulation est beaucoup plus prononcée lorsque la séquence hY3 est dans l'orientation inverse au gène (45hY3.1i) que lorsqu'elle est dans la même orientation (45hY3.1). L'effet de la séquence hY3 inverse se rapproche même de l'effet des sites

de liaison pour la protéine hnRNP A1 (plasmide 36). Donc, l'effet de la séquence hY3 est plus important lorsqu'elle est située en aval de l'exon alternatif, tandis que la séquence hY3 inverse a une influence plus grande lorsque située en amont de l'exon alternatif. Dans tous les cas, la présence de séquences hY3 en amont ou en aval de l'exon alternatif, dans la même orientation que le gène ou dans l'orientation opposée, favorise l'exclusion de cet exon de l'ARN messager mature. Dans ce contexte, les séquences hY3 pourraient être considérées comme des éléments de régulation négative, donc classées dans la catégorie des ISS.

L'action des séquences introniques régulant l'épissage alternatif se fait souvent au niveau de la reconnaissance des sites d'épissage. Dans le modèle d'épissage alternatif hnRNP A1, des changements au niveau de la sélection du site d'épissage 5' influencent les niveaux d'inclusion ou d'exclusion de l'exon 7B. Des séquences régulatrices peuvent favoriser ou inhiber un site d'épissage. Dans le cas où la séquence hY3 est située en aval de l'exon alternatif, le déplacement du choix du site d'épissage observé peut être dû à une inhibition du site 5' proximal ou à une activation du site 5' distal. La distance de la séquence hY3 par rapport au site proximal (30 nt) étant beaucoup plus petite que la distance au site distal (175 nt), il est plus probable que le site 5' proximal soit inhibé. De l'autre côté, la séquence hY3 située en amont de l'exon alternatif est à une distance de 29 nucléotides du site 5' distal et à 121 nucléotides du site 5' proximal. Dans cette situation, l'effet pourrait être dû plutôt à une activation du site d'épissage 5' distal. Donc l'effet semble dépendre du contexte dans lequel les séquences hY sont situées. Par contre, cette analyse, basée exclusivement sur la proximité des sites d'épissage, ne permet pas d'expliquer le

mécanisme par lequel les séquences hY3 modulent l'épissage de l'unité d'épissage alternatif hnRNP A1.

À première vue, la grande variabilité des résultats d'une expérience à une autre, illustrée par des écart-types importants, laisse croire que les effets observés ne sont pas significatifs. Par contre, lorsque analysées séparément, ces expériences montrent un patron d'épissage cohérent. Dans chaque expérience, le contrôle de base 45 est toujours la construction obtenant la plus grande proportion de produit proximal, suivi des constructions 45hY3 (ou 45hY3.1), 45hY3i (ou 45hY3.1i) et 36. Ce sont plutôt les niveaux globaux d'épissage qui changent d'une expérience à une autre. Cette analyse s'applique pour toutes les expériences d'épissage *in vitro* avec le modèle hnRNP A1 effectuées dans cette étude. La variabilité pourrait s'expliquer par des erreurs de manipulation, ou par l'état de l'extrait nucléaire utilisé. Un extrait nucléaire qui a été dégelé puis regelé plus d'une fois peut avoir perdu l'activité de quelques protéines. Dans ce type d'expériences, les patrons d'épissage obtenus sont donc globalement reproductibles.

Afin de confirmer la reproductibilité des résultats, nous avons utilisé 3 différents extraits nucléaires de cellules HeLa, mais préparés indépendamment. Due à l'imprécision de certaines manipulations lors de la préparation des extraits et à l'état des cellules au moment de l'extraction, la composition de ces extraits peut varier. Les patrons d'épissage *in vitro* obtenus avec le modèle hnRNP A1 dans lequel ont été positionnées des séquences hY3 en aval de l'exon alternatif sont semblables pour les 3 extraits nucléaires utilisés : la présence d'une séquence hY3 inverse déplace l'épissage vers le site 5' distal de façon plus marquée que la séquence hY3 orientée

dans le même sens que le gène. Comme mentionné précédemment dans l'explication de la variabilité des résultats d'une expérience à une autre, ce sont plutôt les niveaux globaux d'épissage qui varient d'un extrait nucléaire à un autre. Ces résultats sont donc reproductibles. Cependant, les processus biologiques, et en particulier l'épissage, varient sensiblement d'un type cellulaire à un autre. Pour déterminer si les patrons d'épissage observés avec le modèle hnRNP A1 dépendent du type cellulaire à partir duquel sont préparés les extraits nucléaires, il faudrait répéter les expériences avec des extraits nucléaires préparés à partir d'autres lignées cellulaires que la lignée HeLa.

Dans tous les cas, la séquence hY3 positionnée dans l'orientation inverse à celle du mini-gène déplace l'épissage vers le site 5' distal de façon plus importante que la séquence hY3 orientée dans le même sens que le mini-gène. Il n'a pas été démontré que la séquence hY3 inverse pouvait lier les mêmes protéines que l'ARN hY3, ou quelque autre protéine. Par contre, sa structure secondaire, évaluée par le logiciel *mfold* (MATHEWS et al., 1999; ZUCKER, 2003), adopte la forme d'une tige boucle quasiment aussi stable que celle de l'ARN hY3.

5.3.2 Recherche d'un facteur *trans*

Les séquences régulatrices introniques agissent souvent par la liaison de facteurs d'épissage. Les protéines ainsi recrutées pourraient interférer avec des composantes du spliceosome pour empêcher leur liaison au site d'épissage ou au contraire, favoriser leur association au site d'épissage. Afin de déterminer si la

modulation de l'épissage par des séquences hY3 dépend de l'action de facteurs actifs en *trans*, des expériences dans lesquelles l'extrait nucléaire est préincubé avec différentes quantités d'ARN hY3 (ou hY3 inverse) ont été réalisées. L'ajout d'ARN hY3 compétiteur à la réaction d'épissage du mini-gène hnRNP A1 contenant une séquence hY3 en aval de l'exon alternatif déplace encore légèrement l'épissage vers le site 5' distal. D'une part, la compétition de l'ARN hY3 avec l'ARN pré-messager pour un facteur actif aurait plutôt favorisé un retour de l'épissage vers le site 5' proximal. D'autre part, le même léger effet est observé lorsqu'on ajoute de l'ARN hY3 à la réaction d'épissage du mini-gène hnRNP A1 contrôle (45). De plus, l'ajout d'ARN hY3 inverse dans une réaction d'épissage du mini-gène hnRNP A1 contenant une séquence hY3 inverse en aval de l'exon 7B produit sensiblement le même effet. Donc, ces expériences de compétition montrent que la modulation de l'épissage par une séquence hY3 (et hY3 inverse) positionnée en aval de l'exon alternatif ne semble pas dépendre de l'action d'un facteur en *trans*. Par contre, le déplacement supplémentaire vers le site 5' distal observé dans ces expériences est difficilement explicable et est possiblement non-spécifique. Pour appuyer cette hypothèse, l'ajout d'un ARN quelconque à la réaction d'épissage du contrôle 45 produit sensiblement le même léger effet que l'ajout d'ARN hY3 compétiteur.

Les résultats diffèrent un peu avec les expériences de compétition où une séquence hY3 est située en amont de l'exon alternatif, mais des conclusions semblables peuvent en être tirées. La modulation la plus importante est obtenue lorsqu'une séquence hY3 dans l'orientation opposée au gène est située en amont de l'exon alternatif. L'ajout d'ARN hY3 antisens à cette réaction d'épissage déplace

encore légèrement l'épissage vers le site 5' distal, donc aucune compétition n'est observée. Dans le cas où la séquence hY3 en amont est dans le même sens que le gène, la modulation est très faible comparativement au contrôle. L'ajout d'ARN hY3 ne modifie pratiquement pas la sélection du site d'épissage 5', mais considérant l'effet de cet ARN sur l'épissage du contrôle 45, on peut remarquer une faible compétition. Cependant, comme l'effet sur l'épissage de la séquence hY3 située en amont était plutôt faible, et qu'il s'agit du seul cas où on observe un semblant de compétition, on conclut que la modulation de l'épissage par des séquences hY3 ne dépend d'aucun facteur *trans*.

Ces observations laissent plutôt croire à un mode d'action en *cis*, de nature structurale. Il existe quelques évidences de l'implication de structures secondaires dans l'ARN pré-messager dans la régulation de l'épissage. De telles structures peuvent modifier l'utilisation de sites d'épissage en changeant les distances introniques ainsi qu'en masquant des sites d'épissage, le site de branchement ou d'autres séquences régulatrices (BURATTI & BARALLE, 2004). L'insertion d'une séquence hY3 dans l'intron pourrait changer la structure de l'ARN pré-messager de façon à modifier l'utilisation des sites d'épissage. Les expériences dans lesquelles l'ajout d'une séquence d'ARN complémentaire à la séquence hY3 intronique modifie l'épissage de l'ARN pré-messager en favorisant un retour vers le site d'épissage 5' proximal pourraient supporter cette hypothèse. En effet, les séquences complémentaires, en s'appariant, déstabiliseraient la nouvelle structure de l'ARN pré-messager, ce qui pourrait entraîner un retour de l'épissage vers le site d'épissage habituel. Par contre, cette explication n'est pas valable pour les expériences où les

séquences hY3 sont positionnées en amont de l'exon alternatif, où la modulation n'est pas atténuée par la présence de séquences complémentaires. Dans ce cas, le mécanisme d'action pourrait être différent. Donc, l'effet des séquences hY3 sur l'épissage semble être dépendant de leur position par rapport à l'exon modulé.

5.3.3 Modèle d'épissage alternatif Bcl-x

La présence de séquences correspondant à l'ARN hY3 à proximité de sites d'épissage semble moduler l'épissage alternatif du mini-gène hnRNP A1. Cette modulation ne semble pas dépendre de l'action d'un facteur *trans*, mais plutôt d'un effet structural. Par contre, le mode d'action détaillé reste inconnu. Afin de vérifier s'il s'agit d'un mécanisme général ou d'une particularité du système A1, l'influence des séquences hY3 sur l'épissage a été investiguée à l'aide d'un autre modèle d'épissage alternatif, celui-ci dérivé du gène Bcl-x. D'une part, la grande taille de l'intron de ce modèle permet d'insérer des séquences à de plus grandes distances du site 5', donc d'éliminer les effets potentiellement non spécifiques reliés à la trop grande proximité du site d'épissage et de mieux refléter la réalité génomique (seulement 3 séquences hY3 sont situés à moins de 50nt d'un site d'épissage dans le génome humain). D'autre part, la stratégie de clonage des séquences hY3 dans cet intron permet de conserver relativement intacte la taille de l'intron, donc d'éliminer des effets non spécifiques reliés au changement de taille de l'intron.

L'épissage *in vitro* de cette unité se fait majoritairement au site 5' proximal, donnant l'isoforme Bcl-xL dans plus de 90% des cas. L'insertion d'une séquence

hY3 dans l'intron ne modifie pratiquement pas cette proportion, qu'elle soit à une distance d'environ 100 nt du site d'épissage (Bclx3.1) ou à 400 nt (Bclx3.2). De même, la séquence hY3 inverse positionnée dans l'intron aux mêmes distances ne déplace que très légèrement l'épissage vers le site distal. Ces niveaux de variations sont minimales lorsque comparés à ceux obtenus avec le modèle hnRNP A1. Donc, les séquences hY3 introniques influencent peu l'épissage *in vitro* du mini-gène Bcl-x, peu importe leur orientation et leur distance par rapport au site d'épissage.

Toutes les expériences d'épissage réalisées jusqu'à maintenant dans cette étude ont été effectuées *in vitro* dans un extrait nucléaire de cellules HeLa. Une certaine modulation de l'épissage par des séquences hY3 a été observée dans le modèle hnRNP A1, mais aucune modulation significative n'est observée dans le modèle Bcl-x. La composition de l'extrait nucléaire et la réaction d'épissage *in vitro* pourraient ne pas constituer les conditions optimales pour observer la modulation par les séquences hY3. Afin de s'assurer que l'absence d'effet dans ce dernier modèle n'est pas uniquement dû au fait que l'épissage est effectué *in vitro*, des expériences d'épissage *in vivo* ont été effectuées. Les 2 lignées cellulaires utilisées donnent des résultats semblables, avec des variations de moins de 10% d'une lignée à l'autre et d'une construction à une autre. Encore une fois, ces variations sont faibles comparativement à celles obtenues avec le modèle hnRNP A1. Donc, les séquences hY3 introniques influencent peu l'épissage *in vivo* du mini-gène Bcl-x, peu importe leur orientation, leur distance par rapport au site d'épissage et la lignée cellulaire utilisée.

En conclusion, les séquences correspondant aux ARNs hY situées dans des introns pourraient moduler l'épissage des exons adjacents, comme démontré dans le modèle hnRNP A1. Les expériences de compétition montrent que cette modulation semble se faire plutôt via un réarrangement structural de l'ARN pré-messager. Par contre, les résultats obtenus avec le modèle Bcl-x indiquent que la simple présence de ces séquences hY3 à proximité d'un site n'est pas suffisante pour en modifier significativement l'épissage. Il semble donc qu'un contexte particulier soit nécessaire. Donc, parmi les nombreuses séquences hY3 introniques localisées dans le génome humain, seulement quelques-unes pourraient avoir un rôle dans l'épissage.

REMERCIEMENTS

Je remercie le Dr Gilles Boire, mon directeur de recherche, pour m'avoir permis d'effectuer ma maîtrise dans son laboratoire, pour m'avoir laissé de la latitude et pour avoir cru en moi et m'avoir soutenu même dans les moments plus difficiles. Je remercie les Drs Raymund Wellinger et Sherif Abou Elela, mes co-directeurs de recherche, pour leur implication continue et leurs conseils avisés. J'aimerais remercier le Dr Benoit Chabot, mon co-directeur adoptif, pour son aide précieuse et pour m'avoir fait une place dans son laboratoire dans le projet hY et épissage. Je tiens à remercier les Drs Jean-Pierre Perreault et Benoit Cousineau pour leur participation et leurs conseils dans le projet pseudogènes hY.

Merci à tous les membres du labo Wellinger : Martin Toussaint, Michel Larrivée, Julie Parenteau, Lucie Laliberté, Nancy Laterreur, Catherine LeBel, Alain Dandjinou, Nancy Lévesque, Mélissa Ricard, David Faucher et Amadou Bah. Certains pour m'avoir initié à la levure, d'autres pour leur appui technique et leurs conseils, d'autres pour les discussions, scientifiques ou non, et d'autres pour les passes précises sur la palette. Merci également aux membres du labo Chabot d'avoir gentiment inculqué à l'étranger que j'étais les rudiments de l'épissage, plus particulièrement Jean-François Fiset, Daniel Garneau, Timothée Revil et Caroline Paradis. Merci aux membres des laboratoires des Drs Artur Fernandes et Jean-Luc Parent pour leur soutien technique et pour le partage d'équipements et de produits. Je

remercie également les stagiaires, en particulier Guillaume Tremblay et Jean-François Larrivée qui m'ont donné un bon coup de main.

Merci à Jean-François Lucier pour son aide informatique. Merci à Jonathan Perrault pour son énorme collaboration dans le projet pseudogènes hY et pour « Fortune et Gloire ».

Finalement, je remercie tout spécialement ma famille et mes proches pour leurs encouragements soutenus tout au long de mes études et pour avoir cru en moi. Merci à ma compagne, Karoline Lauzon, la plus importante découverte que j'aie faite dans le cadre de mes études.

RÉFÉRENCES

- Alberts B., Bray D., Lewis J., Raff M., Roberts K., Watson J.D. (1994). *Molecular Biology of the Cell*. Published by Garland Publishing, a member of the Taylor & Francis Group.
- Amir-Ahmady B., Boutz P.L., Markovtsov V., Phillips M.L., Black D.L. (2005). Exon repression by polypyrimidine tract binding protein. *RNA*. 11: 699-716.
- Ausubel, F.M., Brent R., Kinston R.E., Moore D.D., Seidman J.G., Smith J.A., Struhl K. (1993). *Current Protocols in Molecular Biology*. Wiley and Sons inc.
- Ayukawa K., Taniguchi S., Masumoto J., Hashimoto S., Sarvotham H., Hara A., Aoyama T., Sagara J. (2000). La autoantigen is cleaved in the COOH terminus and loses the nuclear localization signal during apoptosis. *J Biol. Chem.* 275: 34465-34470.
- Back S.H., Shin S., Jang S.K. (2002). Polypyrimidine tract-binding proteins are cleaved by caspase-3 during apoptosis. *J. Biol. Chem.* 277: 27200-27209.
- Bateman A., Kickhoefer V. (2003). The TROVE module: a common element in Telomerase, Ro and Vault ribonucleoproteins. *BMC Bioinformatics*. 4: 49.
- Batzer M.A., Deininger P.L. (2002). Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3: 370-379.
- Baudin A., Ozier-Kalogeropoulos O., Denouel A., Lacroute F., Cullin C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 21: 3329-3330.
- Belisova A., Semrad K., Mayer O., Kocian G., Waigmann E., Schroeder R., Steiner G. (2005). RNA chaperone activity of protein components of human Ro RNPs. *RNA*. 11: 1084-1094.
- Ben-Chetrit E., Chan E.K., Sullivan K.F., Tan E.M. (1988). A 52-kD protein is a novel component of the SS-A/Ro antigenic particle. *J Exp. Med.* 167: 1560-1571.
- Ben-Chetrit E., Gandy B.J., Tan E.M., Sullivan K.F. (1989). Isolation and characterization of a cDNA clone encoding the 60-kD component of the human SS-A/Ro ribonucleoprotein autoantigen. *J. Clin. Invest.* 83: 1284-1292.
- Black D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72: 291-336.

Blanchette M., Chabot B. (1999). Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* 18: 1939-1952.

Blencowe B.J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* 25: 106-110.

Boeke J.D., LaCroute F., Fink G.R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* 197: 345-346.

Boire G., Craft J. (1989). Biochemical and immunological heterogeneity of the Ro ribonucleoprotein particles. Analysis with sera specific for the RoY5 particle. *J Clin. Invest.* 84: 270-279.

Boire G., Craft J. (1990). Human Ro ribonucleoprotein particles: characterization of native structure and stable association with the La polypeptide. *J Clin. Invest.* 85: 1182-1190.

Boire G., Gendron M., Monast N., Bastin B., Menard H.A. (1995). Purification of antigenically intact Ro ribonucleoproteins; biochemical and immunological evidence that the 52-kD protein is not a Ro protein. *Clin. Exp. Immunol.* 100: 489-498.

Boise L.H., Gonzalez-Garcia M., Postema C.E. (1993). *bcl-x*, a *bcl-2*-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* 74: 597-608.

Bomsztyk K., Denisenko O., Ostrowski J. (2004). hnRNP K: one protein multiple processes. *Bioessays.* 26: 629-638.

Borer R.A., Lehner C.F., Eppenberger H.M., Nigg E.A. (1989). Major nucleolar proteins shuttle between nucleus and cytoplasm. *Cell* 56: 379-390.

Bouffard P., Barbar E., Briere F., Boire G. (2000). Interaction cloning and characterization of RoBPI, a novel protein binding to human Ro ribonucleoproteins. *RNA.* 6: 66-78.

Bouffard P., Briere F., Wellinger R.J., Boire G. (1999). Identification of ribonucleoprotein (RNP)-specific protein interactions using a yeast RNP interaction trap assay (RITA). *Biotechniques.* 27:790-796.

Boulanger C., Chabot B., Menard H.A., Boire G. (1995). Autoantibodies in human anti-Ro sera specifically recognize deproteinized hY5 Ro RNA. *Clin. Exp. Immunol.* 99: 29-36.

Bouvet P., Diaz J.-J., Kindbeiter K., Madjar J.-J., Amalric F. (1998). Nucleolin interacts with several ribosomal proteins through its RGG domain. *J. Biol. Chem.* 273: 19025-19029.

Brachmann C.B., Davies A., Cost G.J., Caputo E., Li J., Hieter P., Boeke J.D. (1998). Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14: 115–132.

Britten R.J., Kohne D.E. (1968). Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* 161: 529-40.

Brookfield J.F. (2001). Selection on Alu sequences? *Curr. Biol.* 11: R900-901.

Brouha B., Schustak J., Badge R.M., Lutz-Prigge S., Farley A.H., Moran J.V., Kazazian H.H. Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl Acad. Sci. USA*, 100, 5280-5285.

Buratti E., Baralle F.E. (2004). Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.* 24: 10505-10514.

Burd C.G., Dreyfuss G. (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* 13: 1197–1204.

Burge C.B., Tuschl T., Sharp P.A. (1999). Splicing of precursors to mRNAs by the spliceosomes. In Gesteland R., Cech T., and Atkins J. (ed.), *The RNA world*, 2nd ed. Cold Spring Harbor Press, Cold Spring Harbor, N.Y. pp 525-560.

Buzdin A., Gogvadze E., Kovalskaya E., Volchikov P., Ustyugova S., Illarionova A., Fushan A., Vinogradova T., Sverdlov E. (2003). The human genome contains many types of chimeric retrogenes generated through in vivo RNA recombination. *Nucleic Acids Res.* 31: 4385-4390.

Caputi M, Zahler AM. (2001). Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J. Biol. Chem.* 276: 4385043-4385049.

Carstens R.P., Wagner E.J., Garcia-Blanco M.A. (2000). An intronic splicing silencer causes skipping of the IIIb exon of fibroblast growth factor receptor 2 through involvement of polypyrimidine tract binding protein. *Mol. Cell. Biol.* 20: 7388-7400.

Casciola-Rosen L.A., Anhalt G., Rosen A. (1994). Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes. *J Exp. Med.* 179: 1317-1330.

Castelo-Branco P., Furger A., Wollerton M., Smith C., Moreira A., Proudfoot N. (2004). Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol. Cell. Biol.* 24: 4174-4183.

Chabot B. (1994). Synthesis and purification of RNA substrates. Dans : Higgins SJ, Hames BD, eds. *RNA processing: a practical approach*. Oxford; IRL Press, 1: 1-29.

Chakshusmathi G., Kim S.D., Robinson D.A., Wolin S.L. (2003). A La protein requirement for efficient pre-tRNA folding. *EMBO J.* 22: 6562-6572.

Chan E.K., Hamel J.C., Buyon J.P., Tan E.M. (1991). Molecular definition and sequence motifs of the 52-kD component of human SS-A/Ro autoantigen. *J Clin. Invest.* 87: 68-76.

Chan R.C., Black D.L. (1997). The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. *Mol. Cell. Biol.* 17: 4667-4676.

Chang Y.N., Kenan D.J., Keene J.D., Gatignol A., Jeang K.T. (1994). Direct interactions between autoantigen La and human immunodeficiency virus leader RNA. *J Virol.* 68: 7008-7020.

Chen X., Quinn A.M., Wolin S.L. (2000). Ro ribonucleoproteins contribute to the resistance of *Deinococcus radiodurans* to ultraviolet irradiation. *Genes Dev.* 14: 777-782.

Chen X., Smith J.D., Shi H., Yang D.D., Flavell R.A., Wolin S.L. (2003). The Ro autoantigen binds misfolded U2 small nuclear RNAs and assists mammalian cell survival after UV irradiation. *Curr. Biol.* 13: 2206-2211.

Chen, C.D., Kobayashi, R. and Helfman, D.M. (1999) Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev.*, 13: 593-606.

Cheng S.T., Nguyen T.Q., Yang Y.S., Capra J.D., Sontheimer R.D. (1996). Calreticulin binds hYRNA and the 52-kDa polypeptide component of the Ro/SS-A ribonucleoprotein autoantigen. *J Immunol.* 156: 4484-4491.

Chou M.Y., Rooke N., Turck C.W., Black D.L. (1999). hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell. Biol.* 19: 69-77.

Chou M.Y., Underwood J.G., Nikolic J., Luu M.H., Black D.L. (2000). Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. *Mol. Cell.* 5: 949-957.

Cost G.J., Feng Q., Jacquier A., Boeke J.D. (2002). Human L1 element target-primed reverse transcription in vitro. *EMBO J.* 21: 5899-5910.

Costa-Mattioli M., Svitkin Y., Sonenberg N. (2004). La autoantigen is necessary for optimal function of the poliovirus and hepatitis C virus internal ribosome entry site in vivo and in vitro. *Mol. Cell. Biol.* 24: 6861-6870.

Cote C.A., Gautreau D., Denegre J.M., Kress T.L., Terry N.A., Mowry K.L. (1999). A *Xenopus* protein related to hnRNP I has a role in cytoplasmic RNA localization. *Mol. Cell.* 4: 431-437.

Cote J., Simard M.J., Chabot B. (1999). An element in the 5' common exon of the NCAM alternative splicing unit interacts with SR proteins and modulates 5' splice site selection. *Nucleic Acids Res.* 27: 2529-2537.

Craig A.W., Svitkin Y.V., Lee H.S., Belsham G.J., Sonenberg N. (1997). The La autoantigen contains a dimerization domain that is essential for enhancing translation. *Mol. Cell. Biol.* 17: 163-169.

Crosio C., Boyl P.P., Loreni F., Pierandrei-Amaldi P., Amaldi F. (2000). La protein has a positive effect on the translation of TOP mRNAs in vivo. *Nucleic Acids Res.* 28: 2927-2934.

Crouch D., Liebke E.H. (1989). The molecular cloning of a mouse Ro RNA, myl1-like sequence. *Nucleic Acids Res.* 17: 4890.

Dani G.M., Zakian V.A. (1983). Mitotic and meiotic stability of linear plasmids in yeast. *Proc. Natl. Acad. Sci. USA.* 80: 3406-3410.

Deininger P.L., Batzer M.A. (1999). Alu repeats and human disease. *Mol. Genet. Metab.* 67: 183-193.

Dejgaard K., Leffers H., Rasmussen H.H., Madsen P., Kruse T.A., Gesser B., Nielsen H., Celis J.E. (1994). Identification, molecular cloning, expression and chromosome mapping of a family of transformation upregulated hnRNP-K proteins derived by alternative splicing. *J. Mol. Biol.* 236: 33-48.

Del Gatto-Konczak F., Olive M., Gesnel M.C., Breathnach R. (1999). hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. *Mol. Cell. Biol.* 19: 251-260.

Denison R.A., Van Arsdell S.W., Bernstein L.B., Weiner A.M. (1981). Abundant pseudogenes for small nuclear RNAs are dispersed in the human genome. *Proc. Natl. Acad. Sci.* 78: 810-814.

Deutscher S.L., Harley J.B., Keene J.D. (1988). Molecular analysis of the 60-kDa human Ro ribonucleoprotein. *Proc. Natl. Acad. Sci. USA.* 85: 9479-9483.

Dewannieux M., Esnault C., Heidmann T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nature Genet.* 35: 41-48.

Dignam J.D., Lebovitz R.M., Roeder R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* 11: 1475-1489.

Eickbush T.H. (1992). Transposing without ends: the non-LTR retrotransposable elements. *New Biol.* 4: 430-440.

Eperon I.C., Makarova O.V., Mayeda A., Munroe S.H., Caceres J.F., Hayward D.G., Krainer A.R. (2000). Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol. Cell. Biol.* 20: 8303-8318.

Esnault C., Maestre J., Heidmann T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.* 24: 363-367.

Expert-Bezançon A., Le Caer J.P., Marie J. (2002). Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the splicing of the alternative exon 6A from chicken beta-tropomyosin pre-mRNA. *J. Biol. Chem.* 277: 16614-16623.

Fabini G., Rajmakers R., Hayer S., Fouraux M.A., Pruijn G.J., Steiner G. (2001). The heterogeneous nuclear ribonucleoproteins I and K interact with a subset of the ro ribonucleoprotein-associated Y RNAs in vitro and in vivo. *J. Biol. Chem.* 276: 20711-20718.

Fabini G., Rutjes S.A., Zimmermann C., Pruijn G.J., Steiner G. (2000). Analysis of the molecular composition of Ro ribonucleoprotein complexes. Identification of novel Y RNA-binding proteins. *Eur. J. Biochem.* 267: 2778-2789.

Farris A.D., Gross J.K., Hanas J.S., Harley J.B. (1996). Genes for murine Y1 and Y3 Ro RNAs have class 3 RNA polymerase III promoter structures and are unlinked on mouse chromosome 6. *Gene.* 174: 35-42.

Farris A.D., Koelsch G., Pruijn G.J., van Venrooij W.J., Harley J.B. (1999). Conserved features of Y RNAs revealed by automated phylogenetic secondary structure analysis. *Nucleic Acids Res.* 27: 1070-1078.

Farris A.D., O'Brien C.A., Harley J.B. (1995). Y3 is the most conserved small RNA component of Ro ribonucleoprotein complexes in vertebrate species. *Gene.* 154: 193-198.

Faye G., Fukuhara H., Grandchamp C., Lazowska J., Michel F., Casey J., Getz G.S., Locker J., Rabinowitz M., Bolotin-Fukuhara M., Coen D., Deutsch J., Dujon B., Netter P., Slonimski P.P. (1973). Mitochondrial nucleic acids in the petite colonie mutants: deletions and repetition of genes. *Biochimie*. 55: 779-92.

Feng Q., Moran J.V., Kazazian H.H. Jr., Boeke J.D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87: 905-916.

Förch P., Puig O., Kedersha N., Martinez C., Granneman S., Seraphin B., Anderson P. and Valcarcel J. (2000). The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Mol. Cell*. 6: 1089–1098.

Fouraux M.A., Bouvet P., Verkaart S., van Venrooij W.J., Pruijn G.J. (2002). Nucleolin associates with a subset of the human Ro ribonucleoprotein complexes. *J Mol. Biol.* 320: 475-488.

Frank M.B. (1999). Characterization of DNA binding properties and sequence specificity of the human 52 kDa Ro/SS-A (Ro52) zinc finger protein. *Biochem. Biophys. Res. Commun.* 259: 665-670.

Frank M.B., McCubbin V.R., Heldermon C. (1995). Expression and DNA binding of the human 52 kDa Ro/SSA autoantigen. *Biochem. J.* 305: 359-362.

Gallego M.E., Gattoni R., Stevenin J., Marie J., Expert-Bezancon A. (1997). The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the beta-tropomyosin alternative exon 6A. *EMBO J.* 16: 1772-1784.

Garcia-Blanco M.A., Jamison S.F., Sharp P.A. (1989). Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev.* 3: 1874-1886.

Gendron M., Roberge D., Boire G. (2001). Heterogeneity of human Ro ribonucleoproteins (RNPS): nuclear retention of Ro RNPS containing the human hY5 RNA in human and mouse cells. *Clin. Exp. Immunol.* 125: 162-168.

Ghetti A., Pinol-Roma S., Michael W.M., Morandi C., Dreyfuss G. (1992). hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res.* 20: 3671-3678.

Ghisolfi-Nieto L., Joseph G., Puvion-Dutilleul F., Amaric F. Bouvet P. (1996). Nucleolin is a sequence specific RNA-binding protein—characterization of targets on pre-ribosomal RNA. *J. Mol. Biol.* 260: 34-53.

- Gietz R.D., Woods R.A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Meth. Enzymol.* 350: 87-96.
- Ginisty H., Sicard H., Roger B., Bouvet P. (1999). Structure and functions of nucleolin. *J Cell Sci.* 112: 761-772.
- Goldring E.S., Grossman L.I., Krupnick D., Cryer D.R., Marmur J. (1970). The petite mutation in yeast. Loss of mitochondrial deoxyribonucleic acid during induction of petites with ethidium bromide. *J. Mol. Biol.* 52: 323-335.
- Goodier J.L., Ostertag E.M., Kazazian H.H. Jr. (2000). Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* 9: 653-657.
- Goodier J.L., Ostertag, E.M., Du, K., Kazazian, H.H., Jr. (2001). A novel active L1 retrotransposon subfamily in the mouse. *Genome Res.* 11: 1677-1685.
- Goodman M., Porter C.A., Czelusniak J., Page S.L., Schneider H., Shoshani J., Gunnell G., Groves C.P. (1998). Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* 9: 585-598.
- Gottlieb E., Steitz J.A. (1989). The RNA binding protein La influences both the accuracy and the efficiency of RNA polymerase III transcription in vitro. *EMBO J.* 8: 841-850.
- Graveley B.R. (2000). Sorting out the complexity of SR protein functions. *RNA.* 6: 1197-1211.
- Graveley B.R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17: 100-107.
- Graveley B.R., Hertel K.J., Maniatis T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA.* 7: 806-818.
- Green C.D., Long K.S., Shi H., Wolin S.L. (1998). Binding of the 60-kDa Ro autoantigen to Y RNAs: evidence for recognition in the major groove of a conserved helix. *RNA.* 4: 750-765.
- Grimm C., Lund E., Dahlberg J.E. (1997). In vivo selection of RNAs that localize in the nucleus. *EMBO J.* 16: 793-806.
- Grover D., Mukerji M., Bhatnagar P., Kannan K., Brahmachari S.K. (2004). Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition. *Bioinformatics* 20: 813-817.

Harris N.L., Senapathy P. (1990). Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis. *Nucleic Acids Res.* 18: 3015-3019.

Hastings M.L., Wilson C.M. Munroe S.H. (2001). A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA. *RNA.* 7: 859-874.

Heise T., Guidotti L.G., Chisari F.V. (1999). La autoantigen specifically recognizes a predicted stem-loop in hepatitis B virus RNA. *J Virol.* 73: 5767-5776.

Hendrick J.P., Wolin S.L., Rinke J., Lerner M.R., Steitz J.A. (1981). Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. *Mol. Cell. Biol.* 1: 1138-1149.

Herskowitz I. (1988). Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol Rev.* 52: 536-553.

Holcik M., Korneluk R.G. (2000). Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosome entry site element: role of La autoantigen in XIAP translation. *Mol. Cell. Biol.* 20: 4648-4657.

Horke S., Reumann K., Schulze C., Grosse F., Heise T. (2004a). The La motif and the RNA recognition motifs of human La autoantigen contribute individually to RNA recognition and subcellular localization. *J Biol. Chem.* 279: 50302-50309.

Horke S., Reumann K., Schweizer M., Will H., Heise T. (2004b). Nuclear trafficking of La protein depends on a newly identified nucleolar localization signal and the ability to bind RNA. *J Biol. Chem.* 279: 26563-26570.

Houck C.M., Rinehart F.P., Schmid C.W. (1979). A ubiquitous family of repeated DNA sequences in the human genome. *J Mol. Biol.* 132: 289-306.

Huberman J.A., Spotila L.D., Nawotka K.A., el-Assouli S.M., Davis L.R. (1987). The in vivo replication origin of the yeast 2 microns plasmid. *Cell*, 51: 473-481.

Huhn P., Pruijn G.J., van Venrooij W.J., Bachmann M. (1997). Characterization of the autoantigen La (SS-B) as a dsRNA unwinding enzyme. *Nucleic Acids Res.* 25: 410-416.

Ishii T., Ohnuma K., Murakami A., Takasawa N., Yamochi T., Iwata S., Uchiyama M., Dang N.H., Tanaka H., Morimoto C. (2003). SS-A/Ro52, an autoantigen involved in CD28-mediated IL-2 production. *J Immunol.* 170: 3653-3661.

Itoh Y., Rader M.D., Reichlin M. (1990). Heterogeneity of the Ro/SSA antigen and autoanti-Ro/SSA response: evidence of the four antigenically distinct forms. *Clin. Exp. Immunol.* 81: 45-51.

Itoh Y., Reichlin M. (1991). Ro/SS-A antigen in human platelets. Different distributions of the isoforms of Ro/SS-A protein and the Ro/SS-A-binding RNA. *Arthritis Rheum.* 34: 888-893.

Izumi R.E., Valdez B., Banerjee R., Srivastava M., Dasgupta A. (2001). Nucleolin stimulates viral internal ribosome entry site-mediated translation. *Virus Res.* 76: 17-29.

Jacquet S., Mereau A., Bilodeau P.S., Damier L., Stoltzfus C.M., Branlant C. (2001). A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. *J. Biol. Chem.* 276: 40464-40475.

Jasinska A., Krzyzosiak W.J. (2004). Repetitive sequences that shape the human transcriptome. *FEBS Lett.* 567: 136-141.

Johnson J.M., Castle J., Garrett-Engele P., Kan Z., Loerch P.M., Armour C.D., Santos R., Schadt E.E., Stoughton R., Shoemaker D.D. (2005). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302: 2141-2144.

Johnson S., Michalak M., Opas M., Eggleton P. (2001). The ins and outs of calreticulin: from the ER lumen to the extracellular space. *Trends Cell. Biol.* 11: 122-129.

Jurica M.S., Moore M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Molec. Cell* 12: 5-14.

Jurka J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl Acad. Sci. USA.* 94: 1872-1877.

Jurka J., Smith T.F., Labuda D. (1988). Small cytoplasmic Ro RNA pseudogene and an Alu repeat in the human alpha-1 globin gene. *Nucleic Acids Res.* 16: 766.

Kajikawa M., Okada N. (2002). LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* 111: 433-44.

Kan J.L., Green M.R. (1999). Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. *Genes Dev.* 13: 462-471.

Kanopka A., Möhlemann O, Akusjärvi G. (1996). Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* 381: 535-538.

Kazazian H.H. Jr., Wong C., Youssoufian H., Scott A.F., Phillips D.G., Antonarakis S.E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332: 164-166.

Kazazian H.H., Jr. (2004). Mobile elements: drivers of genome evolution. *Science* 303: 1626-1632.

Kelekar A., Saitta M.R., Keene J.D. (1994). Molecular composition of Ro small ribonucleoprotein complexes in human cells. Intracellular localization of the 60- and 52-kD proteins. *J Clin. Invest.* 93: 1637-1644.

Kent W.J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12: 656-664.

Kim J.H., Hahm B., Ki Kim Y., Choi M., Jang S.K. (2000). Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol.* 298: 395-405.

Kim Y.K., Back S.H., Rho J., Lee S.H., Jang S.K. (2001). La autoantigen enhances translation of BiP mRNA. *Nucleic Acids Res.* 29: 5009-5016.

Kim Y.K., Hahm B., Jang S.K. (2000). Polypyrimidine tract-binding protein inhibits translation of bip mRNA. *J Mol. Biol.* 304: 119-133.

Kolosha V.O., Martin S.L. (2003). High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). *J Biol. Chem.* 278: 8112-8117.

Korenberg J.R., Rykowski M.C. (1988). Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* 53: 391-400.

Kranz J.E., Holm C. (1990). Cloning by function: an alternative approach for identifying yeast homologs of genes from other organisms. *Proc. Natl. Acad. Sci. USA.* 87: 6629-6633.

Kufel J., Allmang C., Chanfreau G., Petfalski E., Lafontaine D.L., Tollervey D. (2000). Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. *Mol. Cell. Biol.* 20: 5415-5424.

Kumar S., Tamura K., Jakobsen I.B., Nei M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.

Labbe J.C., Jannatipour M., Rokeach L.A. (1995). The *Caenorhabditis elegans* rop-1 gene encodes the homologue of the human 60-kDa Ro autoantigen. *Gene.* 167: 227-231.

- Labbe S., Thiele D.J. (1999). Copper ion inducible and repressible promoter systems in yeast. *Methods Enzymol.* 306: 145-153.
- Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
- Le Gouill C., Parent J-L., Rola-Pleszcynski M., Stankova J. (1994). Analysis of recombinant plasmids by a modified alkaline lysis method. *Anal. Biochem.* 219: 164.
- Lerner M.R., Andrews N.C., Miller G., Steitz J.A. (1981). Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus. *Proc. Natl. Acad. Sci. USA.* 78: 805-809.
- Lev-Maor G., Sorek R., Shomron N., Ast G. (2003). The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science* 300: 1288-1291.
- Lewis B.P., Green R.E., Brenner S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. USA.* 100: 189-192.
- Lim L.P., Sharp P.A. (1998). Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. *Mol. Cell. Biol.* 18: 3900–3906.
- Lin C.H., Patton J.G. (1995). Regulation of alternative 3' splice site selection by constitutive splicing factors. *RNA.* 1: 234-245.
- Liu H., Zhang W., Reed R.B., Liu W., Grabowski P.J. (2002). Mutations in RRM4 uncouple the splicing repression and RNA binding activities of polypyrimidine tract binding protein. *RNA.* 8: 137-149.
- Liu J., He L., Collins I., Ge H., Libutti D., Li J., Egly J.M., Levens D. (2000). The FBP interacting repressor targets TFIIF to inhibit activated transcription. *Mol. Cell.* 5: 331-341.
- Luan D.D., Korman M.H., Jakubczak J.L., Eickbush T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell.* 72: 595-605.
- Mamula M.J., Silverman E.D., Laxer R.M., Bentur L., Isacovics B., Hardin J.A. (1989). Human monoclonal anti-La antibodies. The La protein resides on a subset of Ro particles. *J. Immunol.* 143: 2923-2928.
- Maraia R., Sakulich A.L., Brinkmann E., Green E.D. (1996). Gene encoding human Ro-associated autoantigen Y5 RNA. *Nucleic Acids Res.* 24: 3552-3559.

Maraia R.J., Sasaki-Tozawa N., Driscoll C.T., Green E.D., Darlington G.J. (1994). The human Y4 small cytoplasmic RNA gene is controlled by upstream elements and resides on chromosome 7 with all other hY scRNA genes. *Nucleic Acids Res.* 22: 3045-3052.

Martin S.L., Bushman F.D. (2001). Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.* 21: 467-75.

Matera A.G., Frey M.R., Margelot K., Wolin S.L. (1995). A perinucleolar compartment contains several RNA polymerase III transcripts as well as the polypyrimidine tract-binding protein, hnRNP I. *J Cell. Biol.* 129: 1181-1193.

Mathews D.H., Sabina J., Zuker M., Turner D.H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288: 911-940.

Mathias S.L., Scott A.F., Kazazian H.H. Jr., Boeke J.D., Gabriel A. (1991). Reverse transcriptase encoded by a human transposable element. *Science* 254: 1808-1810.

Matunis M.J., Michael W.M., Dreyfuss G. (1992). Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Mol. Cell. Biol.* 12: 164-171.

McCauliffe D.P., Zappi E., Lieu T.S., Michalak M., Sontheimer R.D., Capra J.D. (1990). A human Ro/SS-A autoantigen is the homologue of calreticulin and is highly homologous with onchocercal RAL-1 antigen and an aplysia "memory molecule". *J Clin. Invest.* 86: 332-335.

McLaren R.S., Caruccio N., Ross J. (1997). Human La protein: a stabilizer of histone mRNA. *Mol. Cell. Biol.* 17: 3028-3036.

Meerovitch K., Svitkin Y.V., Lee H.S., Lejbkowitz F., Kenan D.J., Chan E.K., Agol V.I., Keene J.D., Sonenberg N. (1993). La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate. *J Virol.* 67: 3798-3807.

Meroni G., Diez-Roux G. (2005). TRIM/RBCC, a novel class of 'single protein RING finger' E3 ubiquitin ligases. *Bioessays.* 27: 1147-1157.

Michael W.M., Eder P.S., Dreyfuss G. (1997). The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNPK protein. *EMBO J.* 16: 3587-3598.

Mironov A.A., Fickett J.W., Gelfand M.S. (1999). Frequent alternative splicing of human genes. *Genome Res.* 9: 1288-1293.

Moran J.V., DeBerardinis R.J., Kazazian H.H. Jr. (1999). Exon shuffling by L1 retrotransposition. *Science* 283: 1530-1534.

Morrish T.A., Gilbert N., Myers J.S., Vincent B.J., Stamato T.D., Taccioli G.E., Batzer M.A., Moran J.V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* 31: 159-165.

Mount S.M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.* 10: 459-472.

Nasim F.U., Hutchison S., Cordeau M., Chabot B. (2002). High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *RNA.* 8: 1078-1089.

Nicolaides N.C., Carter K.C., Shell B.K., Papadopoulos N., Vogelstein B., Kinzler K.W. (1995). Genomic organization of the human PMS2 gene family. *Genomics* 30: 195-206.

O'Brien C.A., Margelot K., Wolin S.L. (1993). Xenopus Ro ribonucleoproteins: Members of an evolutionarily conserved class of cytoplasmic ribonucleoproteins. *Proc. Natl. Acad. Sci. USA.* 90: 7250-7254.

O'Brien C.A., Wolin S.L. (1994). A possible role for the 60-kDa Ro autoantigen in a discard pathway for defective 5S rRNA precursors. *Genes and Dev.* 8: 2891-2903.

O'Brien C.A., Harley J.B. (1992). Association of hY4 pseudogenes with Alu repeats and abundance of hY RNA-like sequences in the human genome. *Gene* 116: 285-289.

Ostertag E.M., Goodier J.L., Zhang Y., Kazazian H.H. Jr. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* 73: 1444-1451.

Ostertag E.M., Kazazian H.H. Jr. (2001). Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* 35: 501-538.

Ostrowski J., Schullery D.S., Denisenko O.N., Higaki Y., Watts J., Aebersold R., Stempka L., Gschwendt, Bomsztyk K. (2000). Role of tyrosine phosphorylation in the regulation of the interaction of heterogenous nuclear ribonucleoprotein K with its protein and RNA partners. *J. Biol. Chem.* 275: 3619-3628.

Padgett R.A., Grabowski P.J., Konarska M.M., Seiler S., Sharp P.A. (1986). Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* 55: 1119-1150.

Page-McCaw P.S., Amonlirdviman K., Sharp P.A. (1999). PUF60: A novel U2AF65-related splicing activity. *RNA.* 5: 1548-1560.

Pannone B.K., Xue D., Wolin S.L. (1998). A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *EMBO J.* 17: 7442-7453.

Pavlicek A., Jabbari K., Paces J., Paces V., Hejnar J.V., Bernardi G. (2001). Similar integration but different stability of Alus and LINEs in the human genome. *Gene* 276: 39-45.

Peek R., Pruijn G.J., van der Kemp A.J., van Venrooij W.J. (1993). Subcellular distribution of Ro ribonucleoprotein complexes and their constituents. *J. Cell Sci.* 106: 929-935.

Pellizzoni L., Lotti F., Rutjes S.A., Pierandrei-Amaldi P. (1998). Involvement of the *Xenopus laevis* Ro60 autoantigen in the alternative interaction of La and CNBP proteins with the 5'UTR of L4 ribosomal protein mRNA. *J Mol. Biol.* 281: 593-608.

Perez I., McAfee J.G., Patton J.G. (1997). Multiple RRM domains contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry* 36: 11881-11890.

Pickeral O.K., Makalowski W., Boguski M.S., Boeke J.D. (2000). Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* 10: 411-415.

Pickering B.M., Mitchell S.A., Evans J.R., Willis A.E. (2003). Polypyrimidine tract binding protein and poly r(C) binding protein 1 interact with the BAG-1 IRES and stimulate its activity in vitro and in vivo. *Nucleic Acids Res.* 31: 639-646.

Pilipenko E.V., Viktorova E.G., Guest S.T., Agol V.I., Roos R.P. (2001). Cell-specific proteins regulate viral RNA translation and virus-induced disease. *EMBO J.* 20: 6899-6908.

Poleev A., Hartmann A., Stamm S. (2000). A trans-acting factor, isolated by the three-hybrid system that influences alternative splicing of the amyloid precursor protein minigene. *Eur. J. Biochem.* 267: 4002-4010.

Pourmand N., Blange I., Ringertz N., Pettersson I. (1998). Intracellular localisation of the Ro 52kD auto-antigen in HeLa cells visualised with green fluorescent protein chimeras. *Autoimmunity* 28: 225-233.

Prujn G.J., Simons F.H., van Venrooij W.J. (1997). Intracellular localization and nucleocytoplasmic transport of Ro RNP components. *Eur. J. Cell. Biol.* 74: 123-132.

Pruijn G.J., Slobbe R.L., van Venrooij W.J. (1991). Analysis of protein--RNA interactions within Ro ribonucleoprotein complexes. *Nucleic Acids Res.* 19: 5173-5180.

Pruijn G.J., Wingens P.A., Peters S.L., Thijssen J.P., van Venrooij W.J. (1993). Ro RNP associated Y RNAs are highly conserved among mammals. *Biochim. Biophys. Acta.* 1216: 395-401.

Quentin Y. (1992). Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. *Nucleic Acids Res.* 20: 3397-3401.

Quinn L.M., Dickins R.A., Coombes M., Hime G.R., Bowtell D.D.L., Richardson H. (2003). *Drosophila* Hfp negatively regulates dmyc and stg to inhibit cell proliferation. *Development* 131: 1411-1423.

Rader M.D., O'Brien C., Liu Y.S., Harley J.B., Reichlin M. (1989). Heterogeneity of the Ro/SSA antigen. Different molecular forms in lymphocytes and red blood cells. *J. Clin. Invest.* 83: 1293-1298.

Ray P.S., Das S. (2002). La autoantigen is required for the internal ribosome entry site-mediated translation of Coxsackievirus B3 RNA. *Nucleic Acids Res.* 30: 4500-4508.

Robberson B.L., Cote G.J., Berget S.M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* 10: 84-94.

Roger B., Moisand A., Amalric F., Bouvet P. (2002). Repression of RNA polymerase I transcription by nucleolin is independent of the RNA sequence that is transcribed. *J. Biol. Chem.* 277: 10209-10219.

Rokeach L.A., Haselby J.A., Meilof J.F., Smeenk R.J., Unnasch T.R., Greene B.M., Hoch S.O. (1991). Characterization of the autoantigen calreticulin. *J Immunol.* 147: 3031-3039.

Rooke N., Markovtsov V., Cagavi E., Black D.L. (2003). Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. *Mol. Cell. Biol.* 23: 1874-1884.

Rose M.D., Winston F., Hieter P. (1990). *Methods in Yeast Genetics: A Laboratory Course Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Rosenblum J.S., Pemberton L.F., Bonifaci N., Blobel G. (1998). Nuclear import and the evolution of a multifunctional RNA-binding protein. *J. Cell. Biol.* 143: 887-899.

Roy-Engel A.M., Salem A.H., Oyeniran O.O., Deininger L., Hedges D.J., Kilroy G.E., Batzer M.A., Deininger P.L. (2002). Active Alu element 'A-tails': size does matter *Genome Res.* 12: 1333-1344.

Sambrook J., Fritsch E.F., Maniatis T. (1989). *Molecular cloning: a laboratory manual*. Second Edition. Cold Spring Harbor Laboratory Press. pp1.25-1.28.

Schmucker D., Clemens J.C., Shu H., Worby C.A., Xiao J., Muda M., Dixon J.E., Zipursky S.L. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101: 671-684.

Shapiro M.B., Senapathy P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15: 7155-7174.

Shen H., Kan J.L., Ghigna C., Biamonti G., Green M.R. (2004). A single polypyrimidine tract binding protein (PTB) binding site mediates splicing inhibition at mouse IgM exons M1 and M2. *RNA*. 10: 787-94.

Sherman F., Fink G.R. Hicks J.B. *Methods in Yeast Genetics*. (1986). Cold Spring Harbor, NY Cold Spring Harbor Laboratory.

Shi H., O'Brien C.A., Van Horn D.J., Wolin S.L. (1996). A misfolded form of 5S rRNA is complexed with the Ro and La autoantigens. *RNA*. 2: 769-784.

Shin-Darlak C.Y., Skinner A.M., Turker M.S. (2005). A role for Pms2 in the prevention of tandem CCTT substitutions induced by ultraviolet radiation and oxidative stress. *DNA Repair* 4: 51-57.

Shnyreva M., Schullery D.S., Suzuki H., Higaki Y., Bomsztyk K. (2000). Interaction of two multifunctional proteins. Heterogeneous nuclear ribonucleoprotein K and Y-box-binding protein. *J. Biol. Chem.* 275: 15498-15503.

Simard M., Chabot B. (2002). SRp30c is a repressor of 3' splice site utilization. *Mol. Cell. Biol.* 22: 4001-4010.

Simons F.H., Broers F.J., Van Venrooij W.J., Pruijn G.J. (1996). Characterization of cis-acting signals for nuclear import and retention of the La (SS-B) autoantigen. *Exp. Cell Res.* 224: 224-236.

Simons F.H., Rutjes S.A., van Venrooij W.J., Pruijn G.J. (1996). The interactions with Ro60 and La differentially affect nuclear export of hY1 RNA. *RNA*. 2: 264-273.

Singer M.F. (1982). SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* 28: 433-434.

Singh R., Valcarcel J., Green M.R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* 268: 1173-1176.

Siomi H., Choi M., Siomi M.C., Nussbaum R.L., Dreyfuss G. (1994). Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell* 77: 33-39.

Siomi H., Matunis M.J., Michael W.M., Dreyfuss G. (1993). The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.* 21: 1193-1198.

Slobbe R.L., Pluk W., van Venrooij W.J., Pruijn G.J. (1992). Ro ribonucleoprotein assembly in vitro. Identification of RNA-protein and protein-protein interactions. *J Mol. Biol.* 227: 361-366.

Slobbe R.L., Pruijn G.J., Damen W.G., van der Kemp J.W., van Venrooij W.J. (1991). Detection and occurrence of the 60- and 52-kD Ro (SS-A) antigens and of autoantibodies against these proteins. *Clin. Exp. Immunol.* 86: 99-105.

Sorek R., Ast G., Graur D. (2002). Alu-containing exons are alternatively spliced. *Genome Res.* 12: 1060-1067.

Southby J., Gooding C., Smith C.W. (1999). Polypyrimidine tract binding protein functions as a repressor to regulate alternative splicing of alpha-actinin mutually exclusive exons. *Mol. Cell. Biol.* 19:2699-2711.

Southern E. (1979). Gel electrophoresis of restriction fragments. *Meth. Enzymol.* 68: 152-176.

Srivastava M., Pollard H.B. (1999). Molecular dissection of nucleolin's role in growth and cell proliferation: new insights. *FASEB J.* 13: 1911-1922.

Stefano J.E. (1984). Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts. *Cell.* 36: 145-154.

Symer D.E., Connelly C., Szak S.T., Caputo E.M., Cost G.J., Parmigiani G., Boeke J.D. (2002). Human l1 retrotransposition is associated with genetic instability in vivo. *Cell* 110: 327-338.

Tange T.O., Damgaard C.K., Guth S., Valcarcel J., Kjems J. (2001). The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J.* 20: 5748-5758.

Teunissen S.W., Kruithof M.J., Farris A.D., Harley J.B., Venrooij W.J., Pruijn G.J. (2000). Conserved features of Y RNAs: a comparison of experimentally derived secondary structures. *Nucleic Acids Res.* 28: 610-619.

Tischendorf J.J., Beger C., Korf M., Manns M.P., Kruger M. (2004). Polypyrimidine tract-binding protein (PTB) inhibits Hepatitis C virus internal ribosome entry site (HCV IRES)-mediated translation, but does not affect HCV replication. *Arch. Virol.* 149: 1955-1970.

Topfer F., Gordon T., McCluskey J. (1993). Characterization of the mouse autoantigen La (SS-B). Identification of conserved RNA-binding motifs, a putative ATP binding site and reactivity of recombinant protein with poly(U) and human autoantibodies. *J Immunol.* 150: 3091-3100.

Van Buskirk C., Schüpbach T. (2002). half pint regulates alternative splice site selection in *Drosophila*. *Dev. Cell.* 2: 343–353.

Van Horn D.J., Eisenberg D., O'Brien C.A., Wolin S.L. (1995). *Caenorhabditis elegans* embryos contain only one major species of Ro RNP. *RNA.* 1: 293-303.

Van Horn D.J., Yoo C.J., Xue D., Shi H., Wolin S.L. (1997). The La protein in *Schizosaccharomyces pombe*: a conserved yet dispensable phosphoprotein that functions in tRNA maturation. *RNA.* 3: 1434-1443.

Wagner E.J., Garcia-Blanco M.A. (2001). Polypyrimidine tract binding protein antagonizes exon definition. *Mol. Cell. Biol.* 21: 3281–3288.

Wagner E.J., Garcia-Blanco M.A. (2002). RNAi-mediated PTB depletion leads to enhanced exon definition. *Mol. Cell.* 10: 943-949.

Wang D.R., Buyon J.P., Chan E.K.L. (1996). Cloning and expression of mouse 60 kDa ribonucleoprotein SS-A/Ro. *Mol. Biol. Rep.* 23: 205–210.

Waterston R.H., Lindblad-Toh K., Birney E., Rogers J., Abril J.F., Agarwal P., Agarwala R., Ainscough R., Alexandersson M., An P., Antonarakis S.E., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-562.

Wei W., Gilbert N., Ooi S.L., Lawler J.F., Ostertag E.M., Kazazian H.H., Boeke J.D., Moran J.V. (2001). Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.* 21: 1429-1439.

Weiner A.M., Deininger P.L., Efstratiadis A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55: 631-661.

Wellinger R.J., Wolf A.J., Zakian V.A. (1993). *Saccharomyces* telomeres acquire single-strand TG1-3 tails late in S phase. *Cell.* 72: 51-60.

- Weser S., Bachmann M., Seifart K.H., Meissner W. (2000). Transcription efficiency of human polymerase III genes in vitro does not depend on the RNP-forming autoantigen La. *Nucleic Acids Res.* 28: 3935-3942.
- Whitcomb J.M., Hughes S.H. (1992). Retroviral reverse transcription and integration: progress and problems. *Annu. Rev. Cell. Biol.* 8: 275-306.
- Wolin S.L., Cedervall T. (2002). The La protein. *Annu. Rev. Biochem.* 71: 375-403.
- Wolin S.L., Steitz J.A. (1983). Genes for two small cytoplasmic Ro RNAs are adjacent and appear to be single-copy in the human genome. *Cell* 32: 735-744.
- Wolin S.L., Steitz J.A. (1984). The Ro small cytoplasmic ribonucleoproteins: identification of the antigenic protein and its binding site on the Ro RNAs. *Proc. Natl. Acad. Sci. USA.* 81: 1996-2000.
- Wollerton M.C., Gooding C., Robinson F., Brown E.C., Jackson R., Smith C.W. (2001). Differential alternative splicing activity of isoforms of polypyrimidine tract binding protein (PTB). *RNA.* 7: 819-832.
- Xie J., Lee J.A., Kress T.L., Mowry K.L., Black D.L. (2003). Protein kinase A phosphorylation modulates transport of the polypyrimidine tract-binding protein. *Proc. Natl. Acad. Sci. USA.* 100: 8776-8781.
- Xue D., Robinson D.A., Pannone B.K., Yoo C.J., Wolin S.L. (2000). U snRNP assembly in yeast involves the La protein. *EMBO J.* 19: 1650-1660.
- Xue D., Shi H., Smith J.D., Chen X., Noe D.A., Cedervall T., Yang D.D., Eynon E., Brash D.E., Kashgarian M., Flavell R.A., Wolin S.L. (2003). A lupus-like syndrome develops in mice lacking the Ro 60-kDa protein, a major lupus autoantigen. *Proc. Natl. Acad. Sci. USA.* 100: 7503-7508.
- Yoo C.J., Wolin S.L. (1997). The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. *Cell.* 89: 393-402.
- Zhang Z., Harrison P., Gerstein M. (2002). Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. *Genome Res.* 12: 1466-1482.
- Zhang Z., Harrison P.M., Liu Y., Gerstein M. (2003). Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* 13: 2541-2558.
- Zhang Z., Schwartz S., Wagner L., Miller W. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* 7: 203-214.

Zhu J., Mayeda A., Krainer A.R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol. Cell.* 8: 1351-1361.

Zuker M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31: 3406-3415.